

CALIFORNIA STATE UNIVERSITY SAN MARCOS

THESIS SIGNATURE PAGE

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

IN

BIOLOGY

THESIS TITLE: Phenology and gene expression of *Paralemanea catenata* (Lemaneaceae, Rhodophyta) in
a Southern California stream

AUTHOR: Nicholas Ainslie

DATE OF SUCCESSFUL DEFENSE:

11/30/2016

THE THESIS HAS BEEN ACCEPTED BY THE THESIS COMMITTEE IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF
SCIENCE IN BIOLOGY

Dr. Robert Sheath
THESIS COMMITTEE CHAIR

Robert Sheath 12/15/16
SIGNATURE DATE

Dr. William Kristan
THESIS COMMITTEE MEMBER

Bill Kristan 11/30/16
SIGNATURE DATE

Dr. Betsy Read
THESIS COMMITTEE MEMBER

Betsy Read 11/30/16
SIGNATURE DATE

Phenology and gene expression of *Paralemanea catenata* (Lemaneaceae, Rhodophyta) in a Southern California stream

Nick Ainslie

CSUSM

Master's Thesis 2016

Abstract

Paralemanea is a freshwater red alga genus in the family Lemaneaceae, of which *Paralemanea catenata* is a member. The cues by which it regulates its triphasic sexual life history are not well known, either in terms of environmental stimuli or mechanisms. The *P. catenata* genome has not been sequenced, and RNA sequencing was used to perform the initial investigation. Comparing morphological and ecological measurements (total length, internodal length, spermatangial ring height, number of nodes, and node width; temperature, pH, total dissolved solids, photoperiod, salinity, and conductivity) taken over its growing season, in concert with the RNA sequencing data, suggests photoperiod is the environmental stimulus by which *P. catenata* regulates its life cycle. Involved in this regulation are phototropins and cryptochromes, which are light-sensitive proteins that allow the alga to continuously monitor light levels and respond to this stimulus.

Introduction

Purpose

Freshwater red algae are studied comparatively less intensively than their marine counterparts; marine reds include commercially important varieties and are thus the subject of intense study to increase yields and gain insight into their life cycles (Kunimoto et al. 1999, Mantri et al. 2009, Chan et al. 2012, Choi et al. 2013). That is to say, there is a lower overall volume of research being conducted on freshwater red algae. The life cycles of the freshwater red algae have been studied, and the stages of development are known. However, it is likely that environmental cues affect the timing of stages, and some stages may not be expressed at all if the environmental conditions are not suitable. Stages of development are ultimately regulated by gene expression, which in turn is triggered by environmental stimuli. Neither the environmental stimuli that provide the ecological trigger for development, nor the genes that are triggered to begin development of a life history stage, are not known for freshwater red algae. Members of the Lemnaceae have been studied ecologically in the east and Midwest of the United States, but not extensively in the southwest.

In recent years, next generation sequencing (NGS) has allowed enormous leaps forward in the ability to sequence and study the genetic material of target organisms. Genes that are involved in regulating the life history stages of an organism can be studied using NGS by sequencing all of the genes that are being expressed during that stage, and then evaluating the function of those genes by matching them with genes of known function from other studies, potentially in very different organisms. Finding a gene that is important in sexual development in a mammal also expressed in red algae suggests that the gene might be playing a part in sexual

development, but inferring functional significance by matches with distantly related, physiologically very different organisms is tenuous. This issue is particularly problematic for taxa that are not model organisms, like the Lemnaceae, as the sequences identified and available on public databases typically only are those used for taxonomic purposes, specifically RuBisCo, ribosomal RNA sequences, and some mitochondrial enzymes. However, using NGS on a poorly studied group is a first step in identifying the genetic regulation of its life history, and is the basis for generating hypotheses that can be studied in greater detail in future work.

This study aims to use *Paralemanea cantata*, a member of the Lemnaceae found in the southwestern United States, to generate hypotheses about the ecological and genetic regulation of life history stages in red algae. Ecological relationships with life history stages were studied in a Southern California stream over the growing season through the gathering of samples weekly from a site in eastern San Diego County. Morphological changes that coincide with ecological measures taken at each sampling location will suggest seasonal cues used by the alga to regulate its life cycle. The genes associated with these changes were assessed by extracting RNA, sequencing it using NGS technologies, and then comparing the transcriptome to NCBI databases. Genes that with functions that are plausibly related to the ecological cues associated with life history stages may be involved in the regulation of the alga's lifecycle. The comparison of the genetic, morphological, and ecological data are the basis for generating hypotheses about how ecological cues trigger the genes that regulate red algae life cycles.

Taxonomy

There are more than 800 genera of algae in the fresh waters of North America (Wehr et al. 2015). Included in these are a number of red algae (Rhodophyta) (Prescott 1962, Blum 1994, Guiry & Guiry 2014). The Rhodophyta is disproportionately under represented in freshwater habitats (< 3% of total species; 66 species and 27 genera in North America) (Sheath 1984, Sheath & Sherwood 2011), with most Rhodophyta members being marine in distribution. This group of algae is united by the following characteristics: they are eukaryotes, lack flagella, use floridean starch for photosynthetic storage, utilize phycobiliprotein pigments in phycobilisomes for light capture, and have unstacked thylakoids (Sheath 2003). Differences between the marine and freshwater red algae include range of sizes, with freshwater red species mostly falling in the range of 1-10 cm in length (Sheath 1984), compared to a much larger average size in marine representatives, < 1 – 30cm (Sheath & Hambrook 1990).

Freshwater red algae mostly occupy flowing waters of streams, from the arctic to the tropics (Sheath & Hambrook 1990). The members of the family Lemnaceae typically are localized in cold, fast-moving waters (Atkinson 1890, Thrib & Benson-Evans 1985). These streams contain higher amounts of dissolved gases, including CO₂, which is required for photosynthesis. A member of this family is *Paralemanea* (Kütz.) Vis et Sheath. *Paralemanea* is a member of the kingdom Plantae, phylum Rhodophyta, class Florideophyceae (which were originally named do to their use of floridean starch for storage) and order Batrachospermales (Sheath 2003, Guiry & Guiry 2011).

Life Cycle

Paralemanea has a relatively complex three-stage life history, composed of haploid and diploid stages, but unlike most of the rest of the Plantae, it contains a second diploid stage, the Chantransia stage (Atkinson 1890, Sheath 1984). The life cycle phases of the genus are as follows: the gametophyte (n) stage is a free-living, pseudoparenchymatous tube growing in the benthos of the stream; the tiny carposporophyte ($2n$) which is the next stage, is formed within the gametophyte as an endophyte, and is the product of gamete fusion; the Chantransia ($2n$) is free living and a series of encrusted filaments attached to the stream bottom. The carpospores are formed after the successful fertilization of the carpogonium, forming the carposporophyte. The Chantransia is formed through the germination of the carpospores which are released by the deterioration of the original gametophyte (Filkin & Vis 2004) (Fig.1). The next generation gametophyte is then formed through somatic meiosis and “buds” or grows off the Chantransia, completing the cycle (Fig.1, Fig. 2).

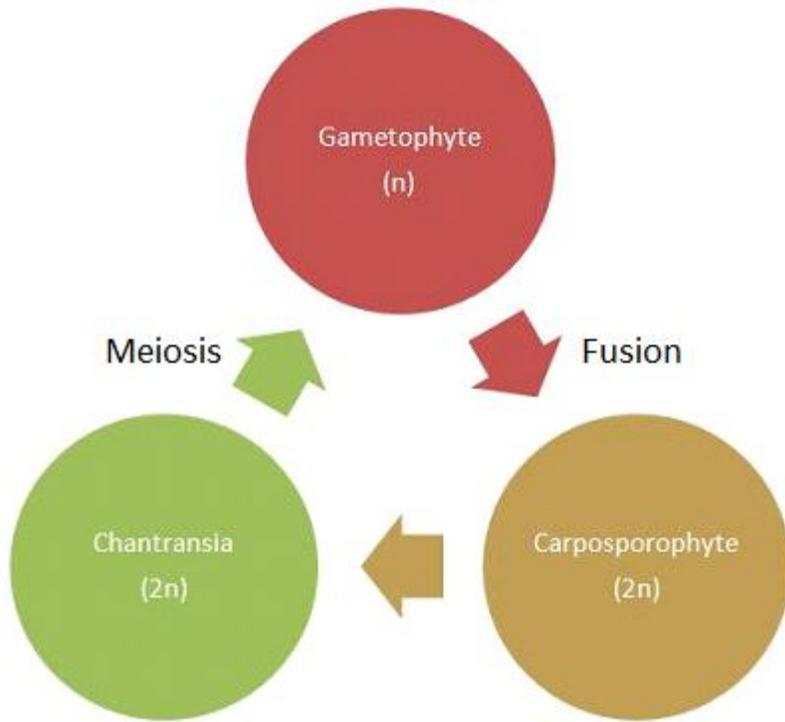


Figure 1 - Life cycle of *Paralemanea*.



Figure 2 – Two views of the Chantransia thallus (2n) “budding” off a haploid cell to eventually form the gametophyte (n) (Indicated by the arrow).

An important difference between the Batrachospermales, including the Lemaneaceae, and most other red algae is the absence of the tetrasporophytic stage. In most other rhodophytes, both the tetrasporophyte and carposporophyte release spores; in the Lemaneaceae only one spore type is released, which is the carpospores (Sheath & Hambrook 1990). This release, coupled with the perennial nature of the Chantransia, serves to make the Lemaneaceae-type life history well adapted to life in streams. In contrast, other red algae which release tetraspores are generally carried downstream by the current in a stream habitat (Graham & Wilcox 2000). This action would result in the population slowly transitioning downstream into

the main trunk of a larger stream or river, which is unfavorable for their growth (Sheath & Sherwood 2011). With the carpospores being released to colonize new areas, and the Chantransia stage maintaining of populations in the upper reaches of stream systems, as proposed by Sheath (1984), *Paralemanea* takes advantage of stream distribution.

Reproductive structures can be observed as the alga proceeds through its life cycle. At the nodes along the gametophyte thallus, spermatangial rings form (Fig. 3); cells of these rings differentiate into the spermatia, non-motile male gametes. The female gametangium, called the carpogonium, is formed in the gametophyte thallus, attached to the carpogonial branch, and the receptive part, the trichogyne, protrudes from the gametophyte thallus. When the spermatangia mature, spermatia are released into the water currents and are probably caught in back eddies where they fertilize the female gamete by fusing with the trichogyne (Fig. 4). The spermatium nucleus fuses with the egg nucleus of the carpogonium to produce the carposporophyte generation that develop and mature in the thallus (Fig. 5). The carpospores are subsequently released when the tips of the thalli degrade and then open. The molecular and seasonal cues by which this cycle is regulated are not well documented (Filkin & Vis 2004).

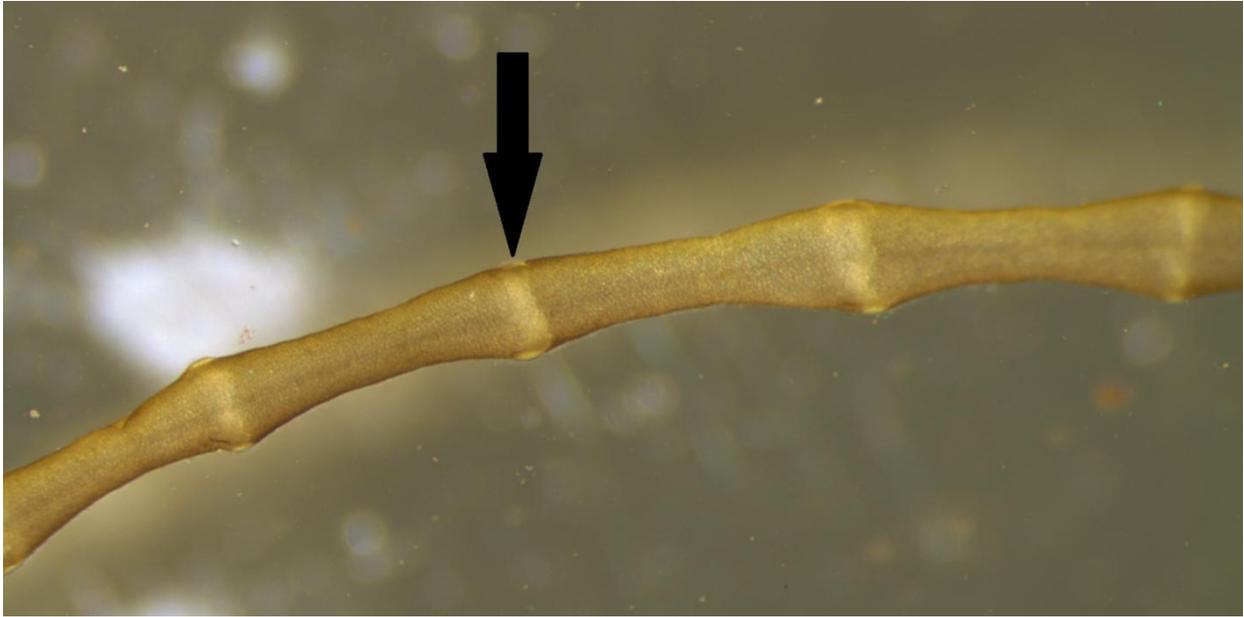


Figure 3 – Spermatangial rings on thallus (arrow).

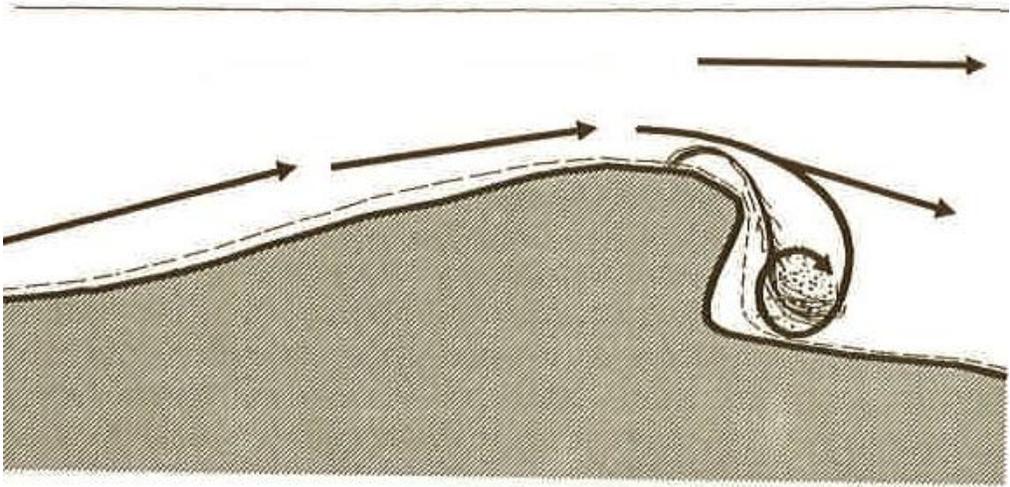


Figure 4 – Stream flow and eddies around benthic algae aid in the fertilization process (from Sheath & Hambrook 1990).

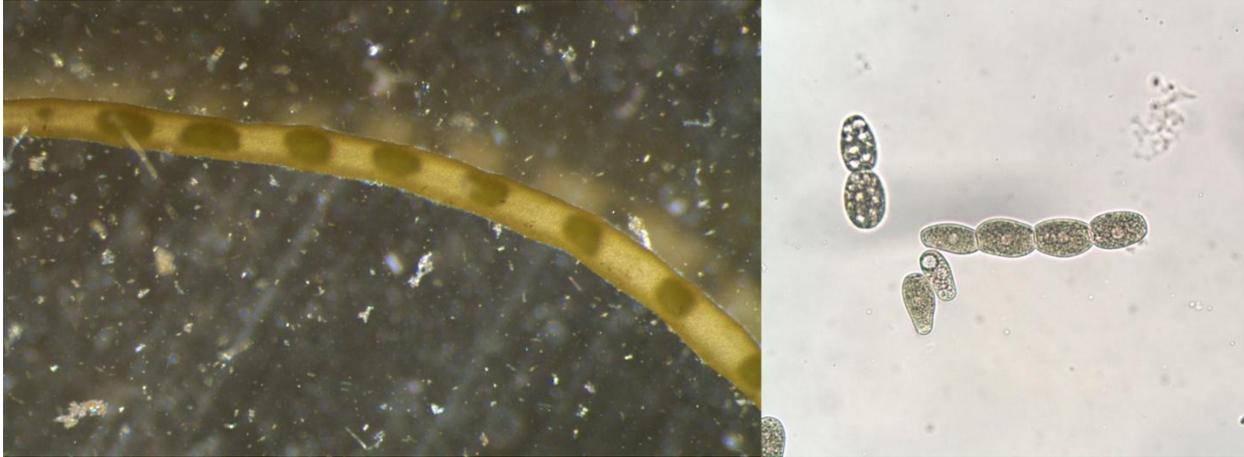


Figure 5 – Carpospores inside the thallus (left) and free floating upon release in short chains (right).

To investigate these molecular and seasonal cues, genetic information must be gathered.

Without a reference genome, RNA is a promising tool to investigate the functional portions of the genome (Flintoft 2011, Grabherr et al. 2011). Other members of the red algae have been investigated using transcriptome analysis to discover genes involved in development and metabolism (Chan et al. 2012), gene regulation in various environmental conditions (Choi et al. 2013), identification of genes of interest for future study such as those relating to stress tolerance and carbon fixation (Yang et al. 2011), and markers for gene discovery and expression profiling (Xie et al. 2013). Other RNA-related studies include the development of a procedure for rapid species ID, using small subunit (SSU) rRNA, when morphological features are not sufficient (Stiller and Waaland 1993), and regional differences between members of the same species using SSU rRNA analysis (Kunimoto et al. 1999).

Transcriptome Sequencing

Using molecular techniques to study an organism without a reference genome can be challenging. When performing transcriptomics in the absence of a sequenced genome, an accurate estimate of genome size is critical to providing meaningful coverage. Without a reference genome, differentiating between coding and non-coding portions of sequenced DNA is problematic; even referencing closely related taxa may not provide sufficient information about specific molecular pathways or gene regulation (Grabherr et al. 2011, Haas et al. 2013). Using RNA bypasses this obstacle since RNA reflects only the functional portions of the DNA. Consequently, transcriptome sequencing is the better choice for the preliminary investigation when little or no sequence information is available. However, the sequences from the genus *Paralemeanea* currently available in the NCBI database are primarily neutral loci that are useful for taxonomy – RuBisCO, 16S, 18S, and 23S ribosomal RNA sequences, and mitochondrial enzymatic sequences; these do not address the questions raised about gene regulation and cues used by this alga to regulate its life cycle. Because of this, this study seeks to develop hypotheses about the genes important in regulation of the *P. catenata* life cycle that can be the basis of future studies.

Materials and Methods:

Field Study

Potential study sites were identified from the Surface Water Ambient Monitoring Program (SWAMP), which samples algae in streams throughout the state. Streams that SWAMP records indicated contained *Paralemeanea*. were checked for the presence of the algae. Sites that were

found to still contain *Paralemanea* were marked on a map, and locations of the populations were identified on the stream bank with brightly colored plastic tape.

Of the sites found to contain *Paralemanea*, one site was used for this study. The site studied was a cold stream with high topographical relief in the eastern San Diego County area, Cold Stream (CS2: 32.94495, -116.56947 at 1331m; Fig. 6). The study ran from March through June, 2013, which encompassed most of the growing season. The whole growing season was not sampled since the population was found after the beginning of the growing season, and the sites identified the previous growing season did not show the presence of this alga that season.

Data were collected from the site weekly. At the location, a stake was planted in the shore to mark the exact location where sampling took place to assure consistency (Sheath & Hambrook 1990, Vis et al. 1991) as water levels and the constitution of the shore could have changed.



Figure 6 – (a) Location of sampling site Cold Stream 2 (CS2) (b) The exact sampling locations

Where *Paralemanea* was present, environmental measurements were taken immediately upstream of the population to avoid stirring sediments. These measurements occurred weekly, with the exception of one missed week in March. Environmental monitoring included measuring pH, dissolved solids (ppm), temperature, conductivity ($\mu\text{S} \cdot \text{cm}^{-1}$), and salinity (ppm S). Total dissolved solids, temperature, conductivity, and salinity were measured with an ExStik II meter (ExTech Instruments Corporation, Nashua, NH); pH was measured with a pHTestr 1

meter (OAKTON Instruments, Vernon Hills, IL). In other studies of freshwater algae current velocity and water chemistry were also analyzed. They were not measured here, as the hardware to measure stream velocity physically was not able to work in the stream sites selected, and the water chemistry was overlooked during the proposal for this study. After ecological measurements were taken, samples of the population were selected for morphological and genetic analysis. The thalli were selected haphazardly and removed with forceps from the stream bottom. To avoid unnecessarily depleting the population, prospective power analysis was performed to ascertain the sample size needed for statistical analysis of relationships between ecological variables and morphological variables. Sample size calculations depend on the statistical method used, and since relationships between numeric variables would be the basis of data analysis, the sample size equation for correlation analysis was used (Zar 1984):

$$N = \left(\frac{z_{\alpha} + z_{\beta}}{C(r)} \right)^2 + 3$$

This equation calculates the needed sample size to achieve the desired power to detect a correlation of a given size. The value from a standard normal distribution for an a level of $\alpha = 0.05$ (z_{α}), a probability of Type II error of $\beta = 0.2$ (which is 1 minus the desired power of 0.8, used to obtain z_{β}), and a Fisher-transformed value of the desired size of correlation to detect ($C(r)$) of 0.3, resulting in a sample size of 36 thalli per sampling instance.

Descriptive Analyses

For each *Paralemanea* specimen the total length, number of nodes, height of the spermatangial rings, internodal length, and width of the nodes were recorded. The presence of carposporophytes was noted when observed. The presence of open-tipped thalli were also noted, since these indicate release of carpospores and decreases thallus length (Sheath 1984).

The microscope used for observation and microphotography of the specimens was the Olympus microscope BX41 with attached Olympus MicroFire S99809 or SC30 digital cameras (Olympus Imaging America Inc., Center Valley, PA).

Statistical Analysis

Statistical analyses were performed with R Version 3.3.1 using the R Commander package. All graphs were generated using Microsoft Excel 2013 (Microsoft Corporation, Redmond, WA, USA). Standard error was used for graphs of algal morphological measurements, and due to the limited data set, and the availability of only one sampling date's complete data, the standard error for the graphs was produced using the following method. First, standard error for each measurement was calculated using the data set available for Mar 10, 2013, and then the percent value relative to the March 10 data point was calculated; finally, the error bars were set as that percentage for all data points. The data from this date was used, as it was the only raw dataset available after recovery was run on the school computer. Associations between ecological and morphological variables were determined using ANOVA and Linear Models. Logistic regression was used to analyze the probability of presence of carpospores and open tips of the gametophytes as a function of environmental variables. Linear regression was used

to determine the relationship of morphological measurements to the environmental variables, as well as to analyze the relationships between morphological measurements.

RNA Analysis

Thalli for RNA analysis were collected once, independently of those used for morphological analysis. Thalli that showed the presence of spermatangial rings and carpospores were chosen, to capture genetic information covering both life cycle stages. To avoid contamination, the *Chantransia* stage was omitted from this analysis, as it was impossible to separate it from epiphytes and sediment.

RNA was extracted using the procedure outlined in Glick (1993). The RNA was then checked for concentration using a spectrophotometer and for purity by running a gel. The extracted RNA was sent to BGI Americas (Cambridge, MA), where a cDNA library was created, and then fragmented for paired-end sequencing. *De novo* sequencing requires much higher quality input data than re-sequencing due to the nature of assembling short reads (Grabherr et al. 2011, Haas et al. 2013). The assembly algorithm is prone to errors due to the smaller data fragments compared to other sequencing endeavors. Raw reads produced by the sequencers were cleaned, then, before *de novo* assembly using the following procedure. Reads with adaptors, low-quality reads (where more than 10% of bases had quality score $Q < 20$), empty reads, and reads with unknown nucleotide sequences $>5\%$ were removed (Grabherr et al. 2011, Haas et al. 2013).

The assembly of the transcriptome from the selected quality reads was undertaken with the *de novo* assembly program Trinity (release 20130225) (Grabherr et al. 2011, Haas et al. 2013).

Paired fragment length and minimum contig length were set to 250 bp and 100 bp, respectively. First, Trinity combined reads having a certain length of overlap into contigs. After the first set of contigs were created, Trinity then mapped all the clean reads back to the contigs. An advantage of paired-end reads is that it is possible to detect contigs originating from the same transcript and the distance between the contigs. The contigs were then connected and assembled into sequences that were not able to be extended on their respective ends. These sequences were then defined as unigenes.

BLASTx alignment between unigenes and protein databases (E-value < 10^{-5}) was then performed. Nt (NCBI nucleotide sequence), Nr (NCBI non-redundant protein), KEGG, Swiss-PROT, and COG databases were queried, the results with the best alignments were used to determine the putative functions of unigene sequences. When conflicting results were returned, information was prioritized as follows: Nr, Swiss-PROT, KEGG, and COG. Unigenes returning no significant hits to any of the databases were run through ESTScan software (Iseli et al. 1999) to predict sequence direction; these were then classified as putative or novel genes.

Functional classification and annotation provides useful insight into the function and expression of unigenes. BLASTx aligned unigene sequences to the GenBank Nr and Swiss-Prot protein databases (E-value < 10^{-5}). In Nr annotation, the Blast2GO (Conesa et al. 2005) software was used to retrieve the GO annotations based on molecular function, cellular component, and biological process ontologies. Once the GO annotations were assigned, WEGO (Ye et al. 2006) software was used to perform GO functional classification and to elucidate the distribution of gene functions from the macro level for all unigenes. Unigene sequences were also queried

against the COG database to predict and classify possible functions. Metabolic pathway analysis was carried out with the KEGG pathway database (Kanehisa & Goto 2000) using BLASTX (E-value < 10^{-5}).

Unigenes

When performing bottom-up RNA sequencing, the reads, which were originally fragmented for sequencing (into short sequence reads – SSRs), must be assembled to perform meaningful analysis. From these SSRs, an assembly program, like Trinity RNA-seq, begins searching for pairs of overlapping reads that can be put together to form longer contiguous reads – a contig (Grabherr et al. 2011). Contigs are then grouped into clusters if they share overlapping segments and have reads spanning the gaps between contigs; from this clustering, the program develops representative de Bruijn graphs of each cluster. De Bruijn graphs are a way to represent all possible subsequences of a certain length from reads in sequencing, and are used to find the most probable assembly to finally form into a transcript. These de Bruijn graphs are evaluated by the final part of the program, which collapses and assembles them into the best-fit transcripts – Unigenes. Another way to think of them is that unigenes are a set of unique transcripts which are believed to have a similar origin point in the genome. These can then be easily queried against NCBI and other databases to annotate their function (Grabherr et al. 2011, Haas et al. 2013).

Results

Field Study

During the sampling period, the stream water temperature ranged from 8.2—14.9°C, from March to June, with the average temperature being 13.2°C. Stream pH varied from 7.4—8.1; it was higher in March, lower in April and May, and then increased again in June (Figs. 7, 8), with an average pH of 7.8. Specific conductance ranged from 265—300 $\mu\text{S} \cdot \text{cm}^{-1}$, with the average being 292 $\mu\text{S} \cdot \text{cm}^{-1}$, and total dissolved solids ranged from 104—212 ppm during the sampling period, with the average being 199.2 ppm. Salinity ranged from 135 in March to 155 ppm S in June (Fig. 8), with an average of 147 ppm S. (Fig. 7) Throughout the sampling period, the plant cover over the stream did not change as it was only covered sparsely near the bank on one side (Fig.6), and there were no noticeable changes in the stream bank at or upstream of the sampling site. Photoperiod data retrieved from NOAA (Fig. 8) only reflects the time between sunrise and sunset. The location of the populations in the stream did not receive direct sunlight this entire time. There is a steep slope to the east, blocking the morning sun, and to the west is a raised road and some boulders, which would shield the location in the afternoon.

Photoperiod data should thus represent a consistent measure of indirect light striking the pool throughout the study.

A number of correlations were significant ($p < 0.05$) among the environmental variables.

Temperature (Temp) was positively correlated with conductivity ($p = 0.003$), total dissolved solids ($p = 0.0012$), salinity ($p = 6.5 \times 10^{-6}$), and photoperiod ($P = 0.0013$). Conductivity was positively correlated with total dissolved solids (ppm) ($p = 2.3 \times 10^{-5}$), salinity ($p = 0.0038$), and photoperiod ($p = 0.041$). Total dissolved solids (ppm) was positively correlated with salinity (ppm

S) ($p=0.0065$). These would indicate that the stream environment changes over the sampling period, which is expected.

A number of morphological variables were significantly related to environmental variables.

Mean Internodal Length (MIL) was associated with photoperiod ($p= 5.7 \times 10^{-5}$), salinity ($p= 0.000198$), and temperature ($p= 0.00326$). Mean Node Width (MNW) was calculated to correlate with pH ($p= 0.021$), photoperiod ($p= 1.59 \times 10^{-5}$), salinity ($p= 0.000196$), and temperature ($p= 0.0101$). Mean Node Number (NN) correlated with temperature ($p= 0.0263$). Mean Total Length (TL) correlated with conductivity ($p= 0.00903$), pH ($p= 0.0298$), photoperiod ($p= 0.000162$), total dissolved solids ($p= 0.0255$), salinity ($p= 1.55 \times 10^{-5}$), and temperature ($p= 0.000385$).

To confirm that the responses to the environmental data were not the product of intra-morphological interactions, these relationships were investigated. Mean spatangial ring height (MSH) was associated with MNW ($p= 4.27 \times 10^{-5}$). TL was associated with all morphological measurements: MIL ($p= 4.64 \times 10^{-7}$), MSH ($p= 0.000314$), MNW ($p= 8.17 \times 10^{-6}$), and NN ($p= 0.00499$). TL also was associated with the presence of carpospores (PC) ($p= 0.000136$). This finding indicates that many of the relationships between the environmental data and morphological measures may be a product of the normal growth of the algae, and not a response relationship.

Between the reproductive events and environmental data, PC was associated with pH ($p= 0.0303$), photoperiod ($p= 0.000127$), salinity (ppm S) ($p= 0.00211$), and temperature ($p=$

0.0355). MSH was associated with photoperiod ($p= 0.0014$) and temperature ($p= 0.0138$). Open tips (OT) of the gametophyte were associated with photoperiod ($p= 0.00742$).

This finding suggests that photoperiod may play a role in determining algal growth and reproduction. In this study, only photoperiod correlated with both the majority of morphological growth metrics and all of the measured reproductive events (mean spermatangial ring height, presence of carpospores, and open tips of the thallus). It follows then that genes involved in light-sensitive pathways may be implicated in the regulation of these processes.

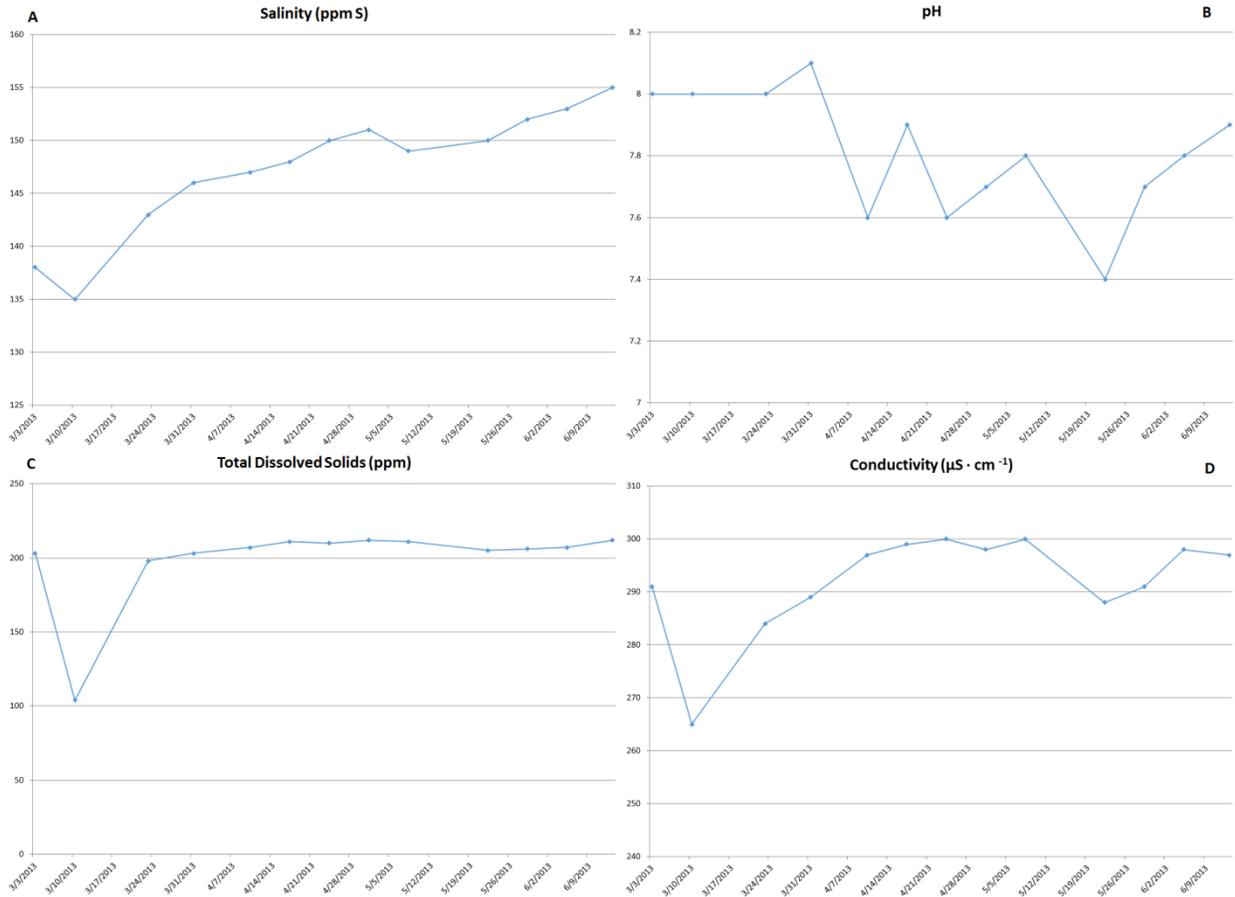


Figure 7 –(A) Salinity (ppm S) (B) pH (C) Total Dissolved Solids (ppm) and (D) Conductivity ($\mu\text{S} \cdot \text{cm}^{-1}$) over the sampling period.

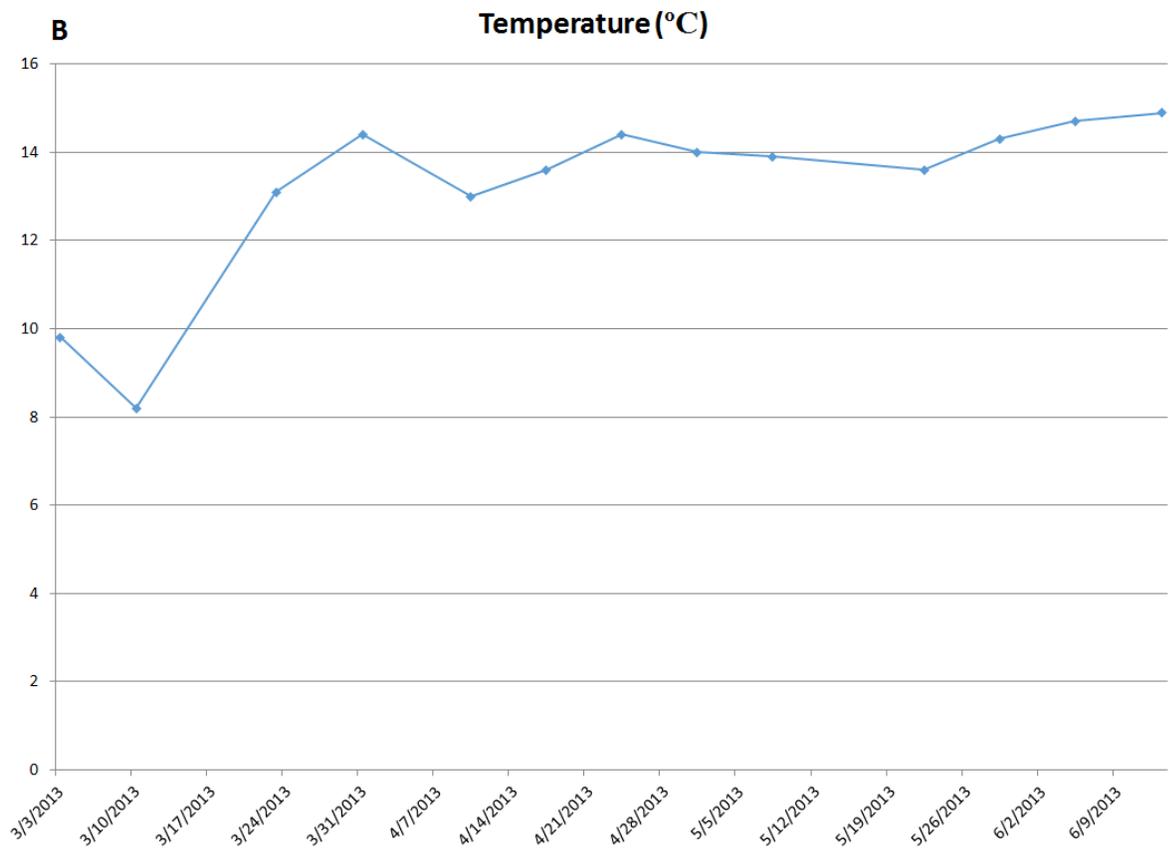
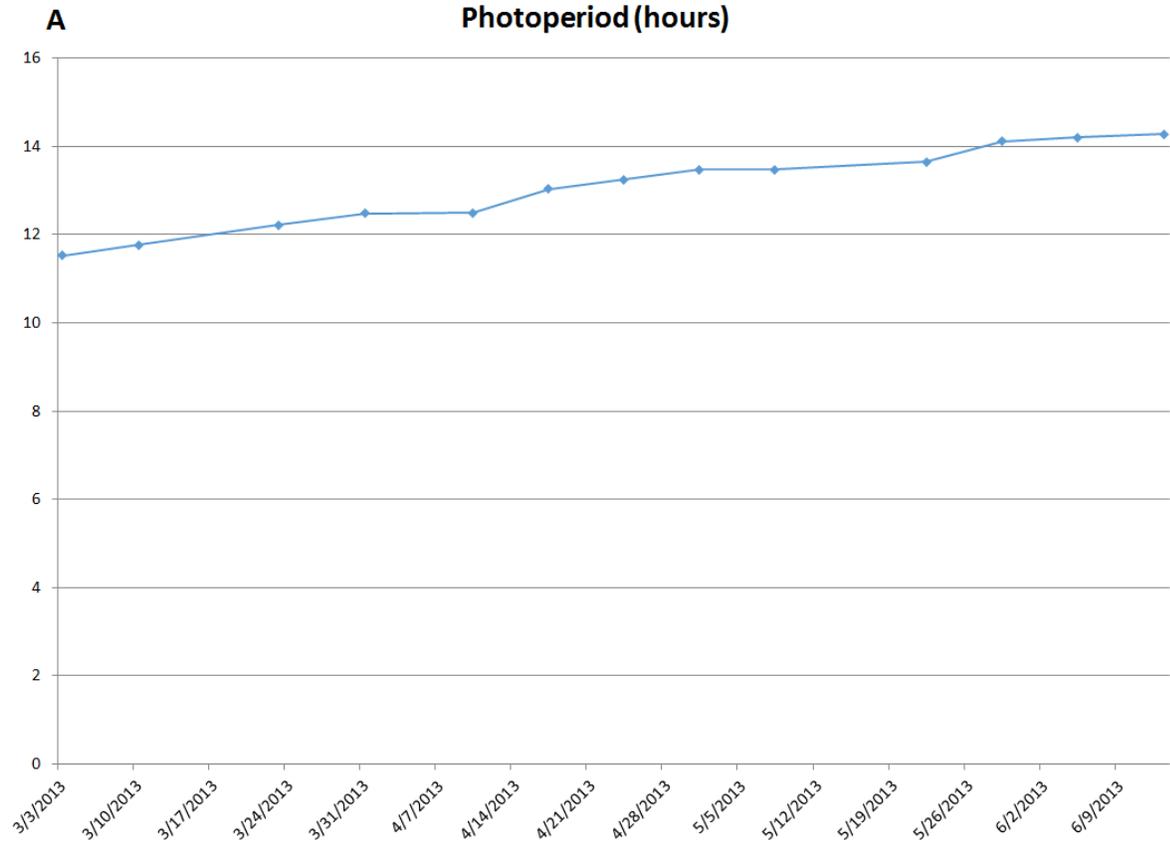


Fig 8 – (A) Photoperiod (hours) and (B) Temperature (°C) over the sampling period.

Gametophytes were present during the entire sampling period, but their mean lengths and widths increased over the sampling period. The total lengths of individuals ranged from 15—99 mm (Fig. 9.0), and the maximum width recorded was 1.29 mm. The number of nodes, mean spatangial ring height, mean node width, and mean internodal length increased over the sampling period (Fig.9.1, 9.2, 9.3, 9.4).

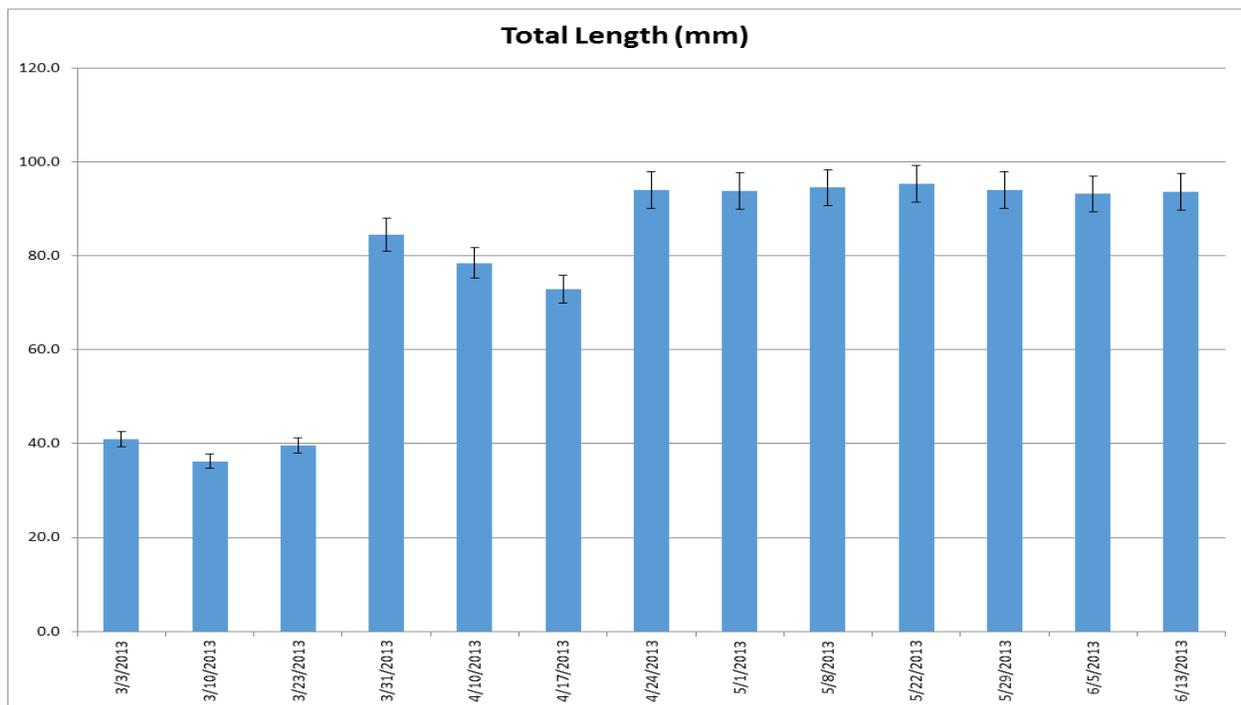


Figure 9.0 – Total Length of the thalli over the sampling period.

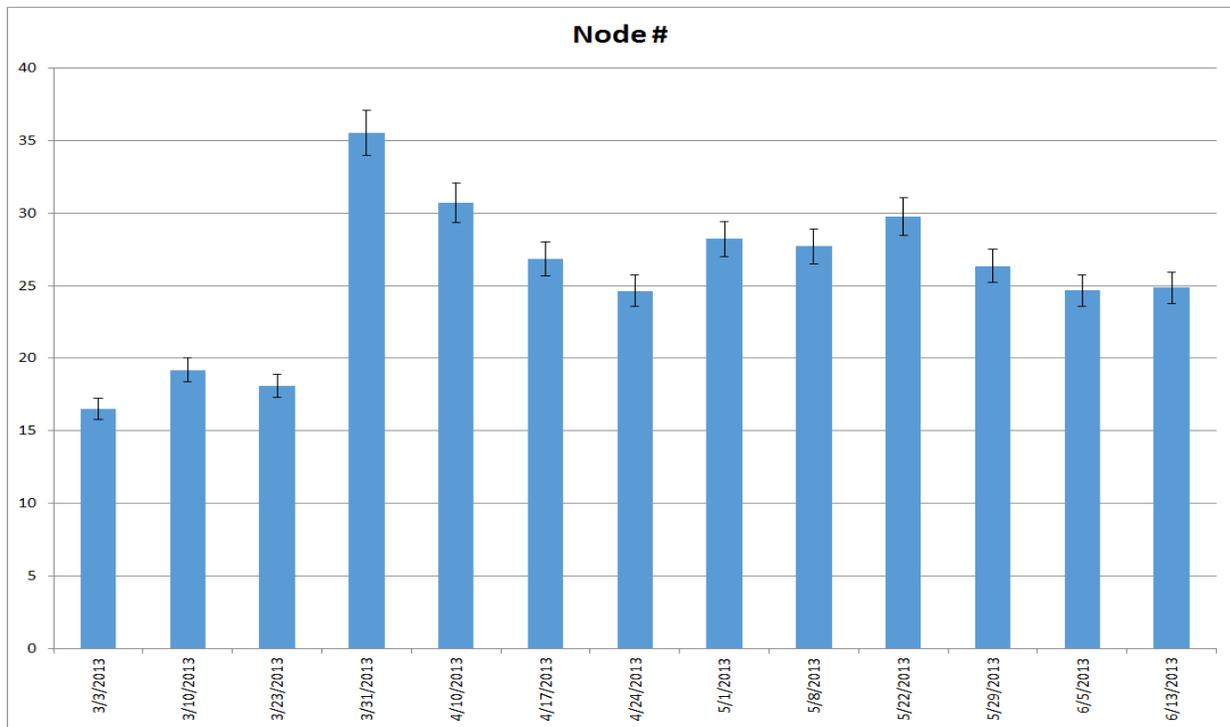


Figure 9.1 – Thr mean number of nodes in the thalli over the sampling period.

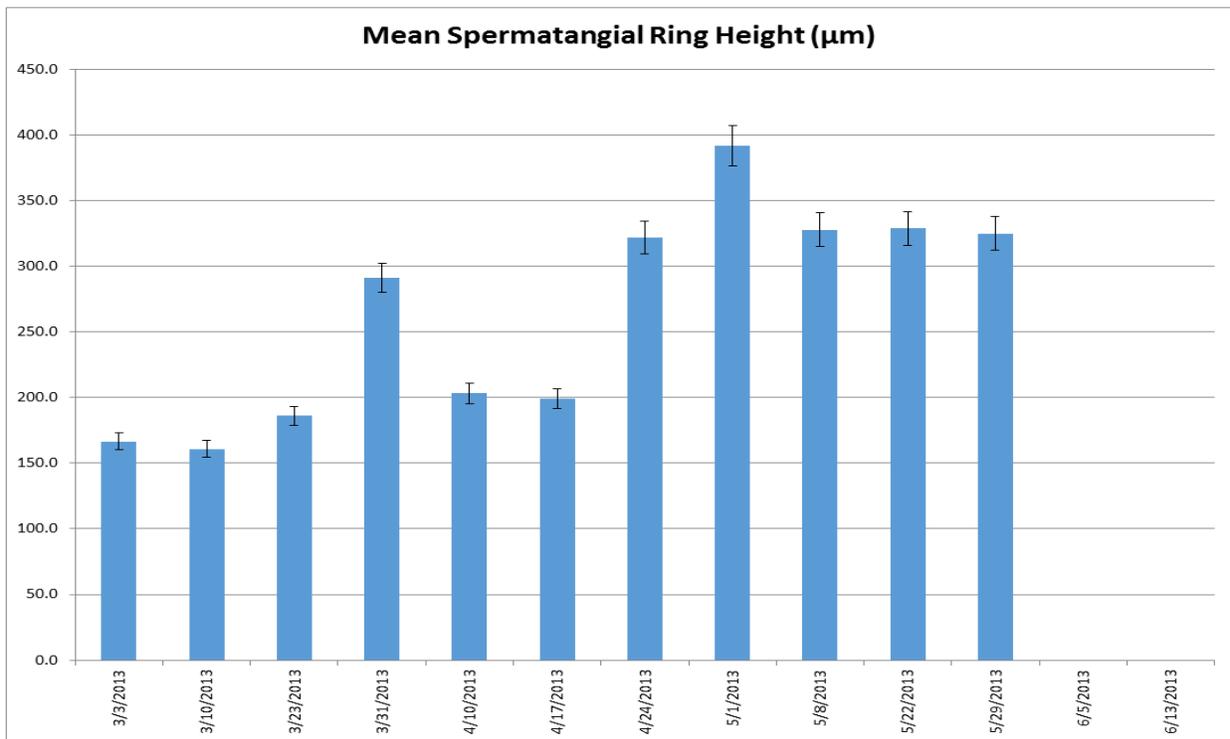


Figure 9.2 – The mean spermatangial ring height over the sampling period.

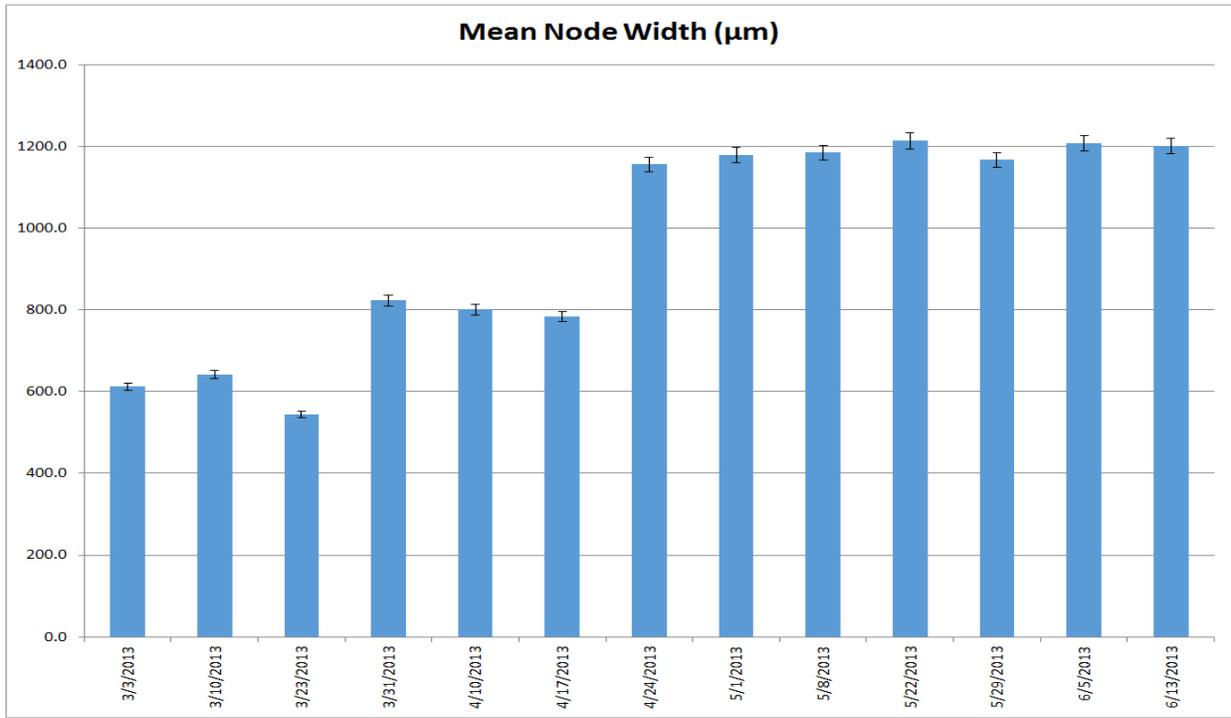


Figure 9.3 – The mean node width over the sampling period.

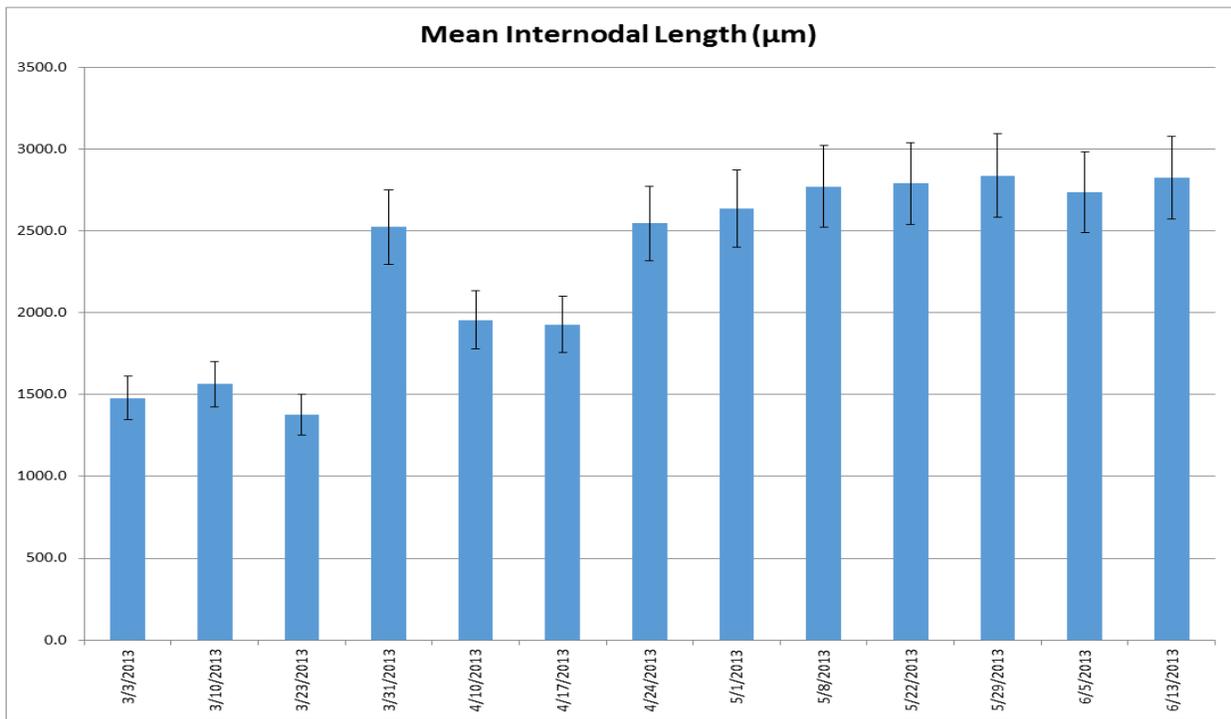


Figure 9.4 – The mean internodal length of the thalli over the sampling period.

Genetic Study

De novo sequencing of the transcriptomes is a valuable tool for investigating the genetics of non-model organisms whose genome sequence are not yet available (Grabherr et al. 2011, Haas et al. 2013). Transcriptome sequencing has arisen as one of the best paths for tracking gene expression and gene discovery (Wang et al. 2009, Flintoft et al. 2011, Grabherr et al. 2011). Next generation sequencing (NGS) technology and assembly programs, such as Trinity, now allow for rapid sequencing and assembly of transcriptomes. Before the advent of these assembly programs, Illumina sequencing was limited to organisms with reference genomes (Wang 2010). To study *P. catenata* in this way, thalli collected May of 2015 showing the presence of both spermatangial rings as well as carpospores were selected for RNA extraction and cDNA library construction. This choice allowed for as many portions of the alga's triphasic life history to be studied simultaneously; but not all, since the chantransia was removed to avoid contamination since it could not be cleaned of sediment or epiphytes. RNA was then extracted using the protocol in Glick (1993), and the extracted RNA was sent to BGI Americas. From this RNA, a cDNA library was constructed, and the resulting libraries were sequenced. A total of 36,537,850 raw reads were generated from the library. After removing low-quality sequences, empty reads, and adaptor sequences, 30,288,506 quality reads were returned (82.9%) with 96.57% Q20 bases. From these quality reads Trinity assembled 130,506 contigs which had an average length of 316 bp (Table1). A further analysis of doing digital counts of unigenes to compare the reads to the unigenes to infer levels of expression would have been useful, but this information was not available in this study, as only the annotated transcriptome was available after the bioinformatics analyses performed by BGI Americas.

Table 1- The Statistics of Assembly Quality.

<u>Term</u>	<u>Value</u>
Total number of raw reads	36,537,850
Total number of clean reads	30,288,506
Q20 score	96.57%
CG Percentage	51.71%
Total number of contigs	130,506
Mean contig length	316
Total number of unigenes	41,970
Mean unigene length	851
N50	1659

The reads were then mapped back to the contigs, which in turn were assembled into unigenes.

From this, 41,970 unigenes with a mean length of 851 bp were generated. Fig. 10 shows the unigene size distribution for this study. The GC content reported was 51.71%.

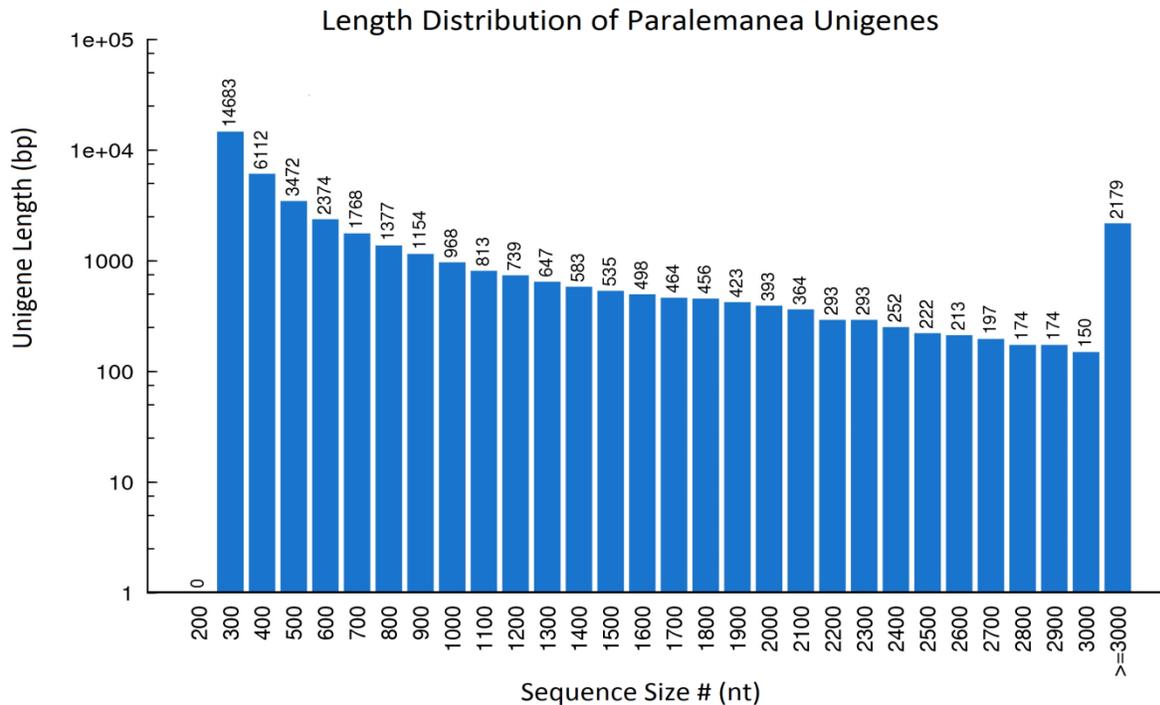


Figure 10 – The length Distribution of *Paralemanea* unigenes.

Functional analysis of the unigenes was accomplished by performing homology searches against the NCBI non-redundant protein database (Nr) and Swiss-Prot protein database with BLASTx (E-value < 10^{-5}). From these searches, 19,514 unigenes (46.5%) returned matches with proteins in the Nr database; 12,602 (30.02%) were found to align to proteins in the Swiss-Prot database (Table 2). Since the genetics of the genus *P. catenata* is relatively unstudied, a large number of unigenes showed no significant homology to sequences the databases queried. In total, 20,908 unigenes (49.81%) returned no significant hits, and may represent novel genes for *Paralemanea*. It should be noted, however, that the majority of organisms in the NCBI databases are model organisms and those related to the study of human-based genetics. Other databases which focus more heavily on non-model organisms and plants would probably have yielded more matches.

The organism in these databases which *P. catenata* shared the most genes with (2612; 13.39%) was *Galdieria sulphuraria*, an extremophile red alga. There was also evidence of some contamination by waterborne pathogens, specifically *Phytophthora megasperma*, *P. infestans*, and *Albugo laibachii*, as they returned high gene homology percentages (5.12%, 2.61%, and 2.74%, respectively). These are members of the oomycote, are common pathogens of algae, and have been studied extensively. The intensity of study means they are over-represented in the NCBI database compared to other groups. Another way to have checked for contamination would be to have plotted the GC content distribution versus unigene numbers to see if there was a normal distribution. If there was a bi-modal or other non-normal distribution, it would

have indicated the presence of non-*Paraleamanea* unigene content. However, in this study, that data was not available, as the bioinformatics analyses were undertaken by BGI Americas.

Table 2 – The transcriptome Annotation Results of *P. catenata*.

<u>Database</u>	Nr	Nt	Swiss-Prot	KEGG	COG	GO	All
<u>Unigenes</u>	19,514	4,817	12,602	14,198	13,599	9,668	21,062

Gene Ontology (GO) represents a standardized classification system for gene function. GO describes the properties of genes and gene products across any organism using a vocabulary and concepts curated by the Gene Ontology project. To undertake GO analysis of the *P. catenata* unigenes, BLAST2GO software was used (Conesa et al. 2005) to generate the GO terms for the unigenes matched in the Nr database. Then, the WEGO program (Ye et al. 2006) undertook GO functional classification of the unigenes. From this analysis, it was determined that 9,668 unigenes returned matches to known proteins and were assigned to three GO classes with 66,740 functional terms (Table 3). Of the three gene ontology classes, most were placed in Biological Processes (29,305; 43.91%), with Cellular Components making the second most populated group (27,173; 40.71%), and finally Molecular Functions (10,262; 15.38%). In the Biological Process category, cell process, metabolic process, and single-organism process constituted most of the category. The most represented classifications in the Cellular Components category were the cell, cell part, and organelle; in the Molecular Function category, catalytic activity, binding, and structural molecule activity were the most highly represented classifications (Fig. 11). Embedded within the Biological Processes class are the Reproduction and Reproductive Processes classifications with 745 and 661 unigenes,

respectively. These two groups only represent about 2.1% of the total functional terms. To determine the significance of the classification of the unigenes into each of these functional term groups, it would be beneficial to normalize these numbers in relation to the size of each database to see whether the number assigned was a function of the number of available terms in each class. This analysis was not possible, as only the output annotation files were available after the bioinformatics analysis by BGI Americas.

Included in the classes mentioned above are the genes coding for light-sensitive proteins called phototropins and cryptochromes, specifically PHOT1, PHOT2, CRY1, and CRYD, which were expressed by *P. catenata*. These proteins are intimately involved in the light-sensitive regulation of many aspects of plant and algal metabolism and physiology (Kianianmomeni & Hallmann 2014), but are not involved directly in the photosynthetic pathway; rather they help the organism optimize certain aspects of its physiology in response to light stimuli. . These are important to note, as photoperiod appears to have the strongest correlation with changes in morphology as well as reproductive processes in *P. catenata*, specifically with the presence of carpospores, open tips of the gametophyte thallus, spermatangial ring height, and total length.

Table 3 – The GO Class Functional Terms.

GO Class	Functional Terms
Biological Processes	29,305
Cellular Components	27,173
Molecular Functions	10,262
Total	66,740

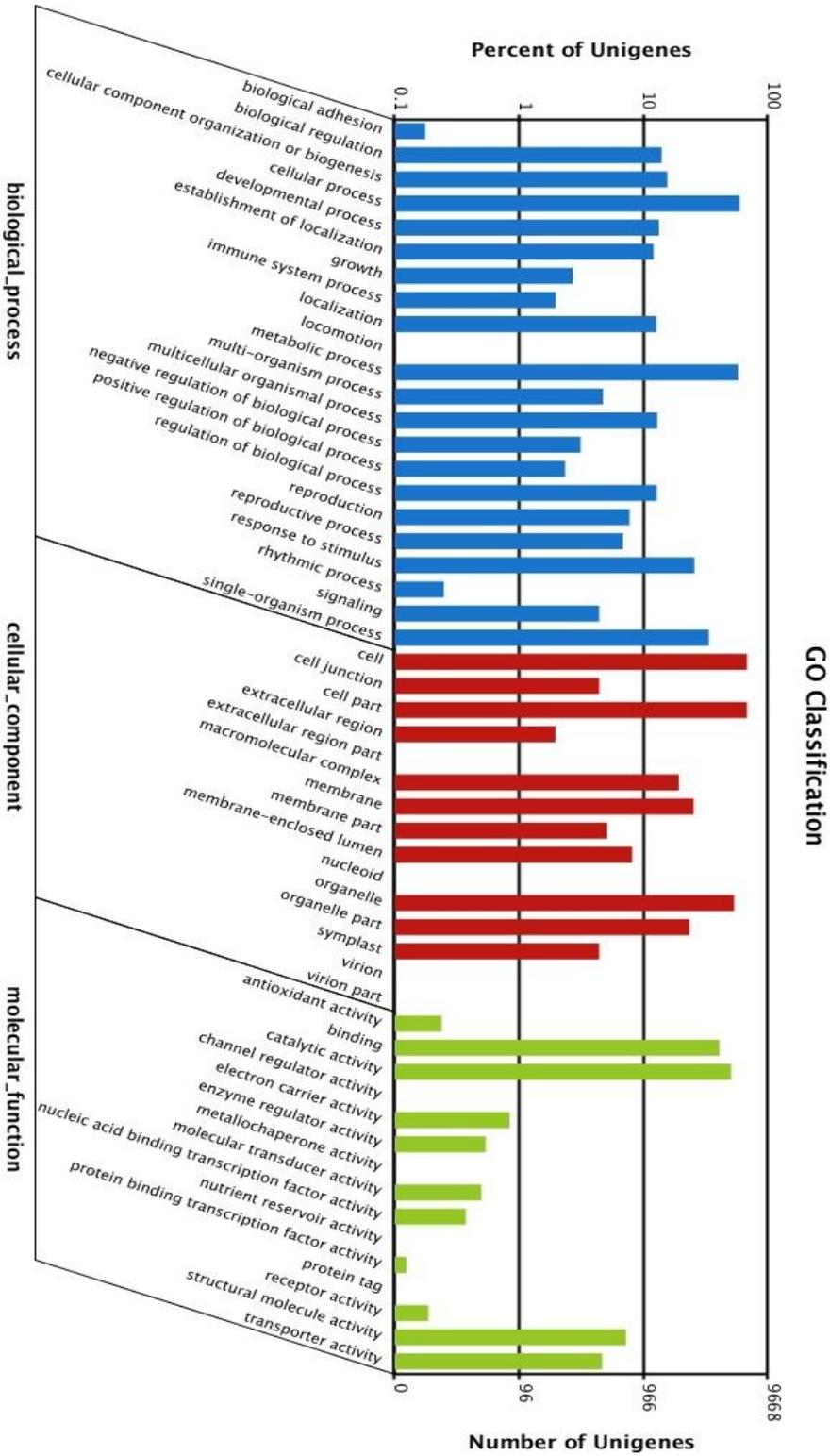


Figure 11 – The GO Class Distribution of unigenes.

For further analysis of the function of unigenes, unigenes were queried against the Clusters of Orthologous Groups (COG) database. This step allows the classification and prediction of the function of the unigenes; this differs from the other analyses in that members of a COG correspond to a shared ancestral domain. From this analysis, 13,599 sequences of 19,514 Nr returns were assigned to one of the 25 COG groups. Of these, aside from the 'General Function Prediction Only' group, Translation, Ribosomal Structure, and Biogenesis constituted the largest category (2,777; 14.23%) (Fig. 12, Table 4). Transcription was the second largest category (2,140; 10.97%), followed by Posttranslational Modification, Protein Turnover, and Chaperones (1855; 9.5%), Carbohydrate Transport and Metabolism (1747; 8.95%), and Replication, Recombination, and Repair (1687; 8.65%), respectively (Fig. 12, Table 4). In three of the five aforementioned classifications, the genes PHOT1, PHOT2, CRYD, and CRY1 are present, specifically Translation, Ribosomal Structure, and Biogenesis, Posttranslational Modification, Protein Turnover, and Chaperones, and Replication, Recombination, and Repair, suggesting their role in these functions.

Figure 12 – The graph of COG Unigene Function Classification.

COG Function Classification

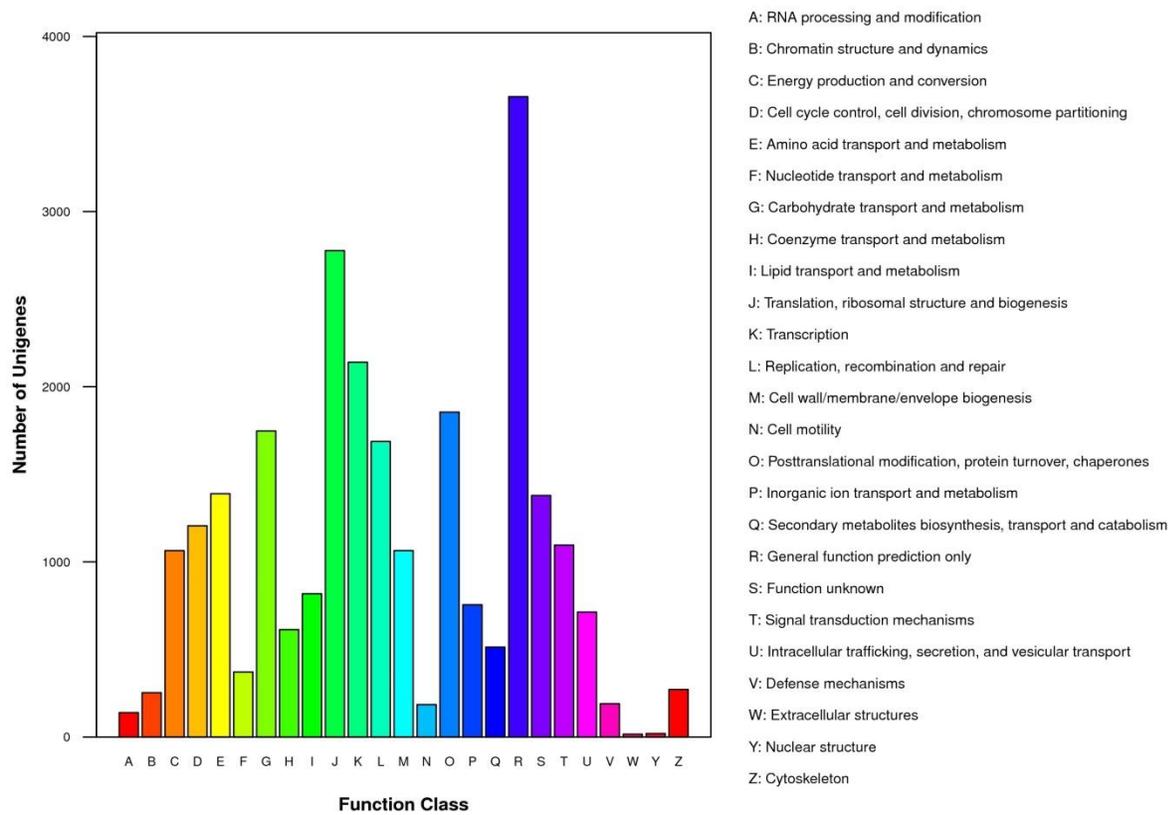


Table 4 – The COG Classification of Unigenes.

COG Classification	Unigenes
Translation, Ribosomal Structure, and Biogenesis	2,777
Transcription	2,140
Posttranslational modification, Protein turnover, and Chaperones	1,855
Carbohydrate Transport and Metabolism	1,747
Replication, Recombination, and Repair	1,687
Total	10,206 (52.3%)

The Kyoto Encyclopedia of Genes and Genomes (KEGG) database is an important tool for understanding molecular interactions among specific organisms (Kanehisa & Goto 2000). This database allows for pathway-based investigations, and helps elucidate genes' interactions and their functions. Using BLASTx (E-value 10^{-5}) to search the database it was found that 14,198 (67.41%) unigenes returned hits from the KEGG database which were then split into 127 pathways. Those designated Metabolic Pathways constituted the highest proportion with 3989 (28.1%) unigenes, followed by Biosynthesis of secondary metabolites with 1613 (11.36%), and then RNA transport with 1351 (9.52%) unigenes (Table 5).

Table 5 - Unigenes Assigned to the Most Common KEGG Pathways.

KEGG Pathway	Unigenes
Metabolic Pathways	3,989
Biosynthesis of Secondary Metabolites	1,613
RNA Transport	1,351
Total	6,953 (33.01%)

Discussion

Field Study

The stream environmental reading of water temperature (8.2—14.9°C), pH (7.4—8.1), and specific conductance (265—300 $\mu\text{S} \cdot \text{cm}^{-1}$) fell within the ranges observed in other studies of the Lemnaceae, with a major difference being the partial growing period being represented in this study versus the whole season in those mentioned hereafter. In Vis & Sheath (1992), the

Lemaneaceae are said to occur in cooler waters averaging 13°C, with specimens collected in streams with pH ranging from 5.0 to 8.6, and specific conductance from 20 to 500 $\mu\text{S} \cdot \text{cm}^{-1}$. In Thrib & Benson-Evans (1985), the mean water temperatures at the sampling stations ranged from 0°C to 16.0°C from November to June and the pH at the sites ranging from 6.9 to 8.7 during the sampling period. In Vis et al. (1991), the water temperature ranged from 2°C in December and January to 12°C in the summer and early autumn, with a pH of 6.2. In Filkin & Vis (2004), the water temperature ranged from 0°C in winter to 22°C in summer, with pH ranging from 6.3 – 7.8, and specific conductance from 110 – 300 $\mu\text{S} \cdot \text{cm}^{-1}$. In Jiménez et al. (2014) water temperature ranged from 14°C – 18°C, pH from 6.7 – 8, and specific conductance ranging from 165 – 248 $\mu\text{S} \cdot \text{cm}^{-1}$. From this information, it appears the Lemaneaceae inhabit waterways with different conditions, with cool water temperature and high current velocity being the common ecological parameters between them all when the literature is reviewed.

Gametophytes were present during the entire sampling period of March through June, 2013, with spermatangial rings readily evident from March until the end of May. In Filkin & Vis (2004) only 20-25% of thalli showed spermatangial rings from February through June when collected, whereas all observed thalli in this study had them concurrently. This may be the result of the differing environments the algae inhabit, as Cold Stream in this study was unshaded where the populations of *Paralemanea* were observed, where there was tree cover in the Ohio woodland stream studied by Filkin & Vis (2004). In addition, due to the ephemeral nature of the populations in Southern California, only part of the growing period was sampled; the

population was discovered after the growing season had already begun. Further study is required to investigate this difference.

Carpospores were first observed at the end of April, and the tips of the thalli were observed as open at the end of May. Similarly, in Vis et al. (1991) and Filkin & Vis (2004), carposporophytes were observed in the late spring/early summer (May through June), with open tips appearing about the same time. In both Vis et al. (1991) and Thrib & Benson-Evans (1985), the members of the Lemnaceae studied were only present for about six months of the year, suddenly disappearing at the beginning of summer. This agrees with the characterization of the end of the growing season presented in this study. Two follow-up visits to the Southern California site in the months after the final sampling instance confirm the disappearance of the gametophyte from the benthos. It was also put forth in Thrib & Benson-Evans (1985) that *Lemanea fluviatilis* was competing with *Cladophora glomerata* and the moss *Rhynchostegium*, and while filamentous green algae were present in the stretches of the stream where *Paralemanea* was absent, it does not appear to be the case in this stream as their coverage did not overlap. Compared to other members of the Lemnaceae, namely *Lemanea fucina* and *P. mexicana* (Vis et al. 1991, Jiménez et al. 2014) which have gametophytes present year-round, *P. catenata* has a limited growing season.

Genetic Study

In this study, 41,970 unigenes with a mean length of 851 bp were generated (Table 1). This mean length was markedly longer than those assembled in other studies of the members of Plantae, including the sweet potato (Solanales) (581 bp) (Wang et al. 2010), *Pyropia yezoensis*

(Rhodophyta) (419 bp) (Yang et al. 2011), peanut (Fabales) (619 bp) (Zhang et al. 2012), and banana (Zingiberales) (554 bp) (Li et al. 2012). The GC content was 51.71%, which is greater than many higher plants, but lower than other algae such as *Chlamydomonas reinhardtii*, which has a GC content of 64% (Merchant et al. 2010). It has been shown that more primitive plant groups have a higher GC content (Von Stackelberg et al. 2006). The differences become apparent when compared to GC contents of other organisms: *Pyropia yezoensis* (63.2%) (Yang et al. 2011), *Adiantum capillus-veneris* (45.97%), rice (47.52%), *Arabidopsis thaliana* (41.1%), *Physcomitrella patens* (47.6%), and *Marchantia polymorpha* (54.75%) (Victoria et al. 2011).

Of particular interest in this study were genes related to reproduction and sexual life cycle of *P. catenata*, specifically those involved in triggering and the regulation of the various stages of its life cycle. These include differentiation of somatic cells into reproductive structures such as spermatangial ring cells differentiating into spermatia and the genesis of the carpogonia and trichogyne, and the changes in the gametophyte thallus needed to complete the reproductive cycle namely the deterioration of the tips of the thallus for the release of carpospores.

It has been mentioned that the specific factors by which *P. catenata* regulates its sexual life cycle are not well known. Many plants and algae take their cues from the environment around them, and it is possible to investigate the genes which are involved in the response to these external stimuli. From the GO database query, 2475 unigenes were returned which explicitly relate to response to stimuli, such as changing photoperiod and salinity. Other classifications which would also relate are the Positive Regulation of Biological Processes (226 unigenes) and

Negative Regulation of Biological Processes (301 unigenes) as these positive and negative feedback pathways may be fed by response to stimuli.

Once the stimuli are received, the next step would be *Paralemanea's* production of reproductive tissue and structures. These steps would be the formation of the spermatangial rings, maturation and release of the spermatia, the formation of the carpogonia and trichogyne, and the formation of carposporophytes and maturation of the carpospores (Sheath & Hambrook 1990). From the COG database query, the classifications Replication, Recombination, and Repair (1687 unigenes), Cell Wall/Membrane/Envelope Genesis (1064 unigenes), and Extracellular Structures (16 unigenes) pertain to this step. The creation of these gametes and spores require the replication of their genetic material, and the creation of specialized membranes, as well as the extracellular structures to hold and subsequently release them (Kugrens & West 1973; Mantri et al. 2009). From the GO query, Cellular Component Organization or Biogenesis (1497 unigenes), Biological Regulation (1349 unigenes), Reproduction (745 unigenes), and Reproductive Process (661 unigenes) are involved here. These steps are part of the reproductive process, and require the regulation of their biological processes for the cue to begin and release, and the genesis of their respective membranes, encapsulating layers, and associated tissues (Mantri et al. 2009).

The final step is the degradation of the thalli, resulting in the open tips through which the carpospores are released. From the COG database query the unigenes in the classification Replication, Recombination, and Repair are thought to play a role, as the degrading thallus may be a result of a deliberate pileup of DNA damage and other damages not being actively repaired

by the cell (Lesser & Farrell 2004; Danon et al. 2006). From the GO database query, the unigenes classified as Biological Regulation, as well as Positive Regulation, Negative Regulation, and Regulation of Biological Processes are thought to play a role. The programmed cell death in response to some kind of cue may be mediated through some discrete molecular pathway, or through some positive or negative feedback pathway (Simon et al. 2000; Lesser & Farrell 2004), such as the up or down regulation of DNA repair or reactive oxygen scavenging enzyme genes.

One common group of genes present throughout all of the aforementioned COG and GO classes were those coding for phototropins (PHOT1 and PHOT 2) as well as cryptochromes (CRY1 and CRY2). Phototropins and cryptochromes are photosensitive proteins which sense light to aid in the acclimation of physiological processes to external stimuli (Foster & Smyth 1980, Kianianmomeni & Hallmann 2014). Recently, with the advent of high-throughput sequencing, a number of algal photoreceptors have been identified and characterized, leading to a better understanding of photoregulated cellular processes and their molecular apparatus (Kianianmomeni & Hallmann 2014). Photosensitive proteins present in algae include phototropins and cryptochromes. Phototropins are present in plants, where cryptochromes are present in both plants and animals (Huang & Beck 2003, Reisdorph & Small 2004). Phototropins and cryptochromes are involved in everything from the positioning and activity of leaves and chloroplasts, stomata opening and closing, to circadian clock regulation, photoperiod-dependent flowering, and seedling growth (Kami et al. 2010, Chaves et al. 2011). Among the morphological measures of growth and maturity in the thallus, as well as reproductive processes, photoperiod may be the cue by which the alga regulates its growth and reproductive

processes. These, along with the evidence from previous studies showing the role of phototropins and cryptochromes in the regulation of many portions of somatic and reproductive cell growth and physiological responses to stimuli, these genes were sought out in *P. catenata*'s RNA transcriptome in this study to test the hypothesis regarding their role in the regulation of its life cycle.

The first major changes in the gametophyte thallus are the development of spermatangial rings, and the release of the mature spermatia to fertilize the carpogonium by attaching to the trichogyne. While the specific genetic mechanisms for this are not known, phototropins may be involved in the triggering of this process. Photoperiod was positively correlated with the morphological measurements relating to this period: Total Length ($p=0.000162$), Mean Node Width ($p=1.59 \times 10^{-5}$), Mean Spermatangial Ring Height ($p=0.0014$), and Mean Internodal Length ($p=5.7 \times 10^{-5}$). In Huan et al. (2002), it was demonstrated that the phototropin CrPhot, in the green alga *Chlamydomonas reinhardtii*, is involved in gametogenesis and the light-induced gamete maturation process. This phototropin performs the functions of both PHOT1 and PHOT2 (Huang et al. 2002), which are the two phototropins present in higher plants, such as *Arabidopsis thaliana*. This same phototropin regulates the ability to successfully reproduce (mating competence) (Huang & Beck 2003). As far as the cues for the development of the reproductive structures, cryptochromes are involved in the light-dependent regulation of the formation of reproductive structures in higher plants (Chaves et al. 2011).

As for the fusion of the spermatia to the trichogyne, and subsequent dissolution of the cell wall between them allowing the nucleus of the spermatium to travel down to the carpogonia,

phototropins may play a role here, as well. CrPhot affects the transcript levels of a cell-wall degrading enzyme, gamete lytic enzyme, which leads to disintegration of target cells or cell structures (Huang & Beck 2003), which may play a role in this process. These cryptochromes and CrPhot's rhodophyte forms (from *Galdieria sulphuraria*, CRY1, PHOT1 & PHOT2) were found in the sequenced genetic data, where CRY1 was classified as involved with Growth and Reproduction, and PHOT 1 and PHOT2 classified as involved with Regulation of Biological Processes and Reproduction in the GO classifications. 12 unigenes were associated with the phototropins and 7 unigenes associated with the cryptochromes (Table 6). Since phototropins and cryptochromes are highly conserved across plants and algae (Onodera et al. 2005), they may perform similar functions in *P. catenata*

Table 6 - Phototropin and Cryptochrome Unigenes found in the transcriptome of *P. catenata*.

Phototropins	12
Cryptochromes	7
DASH Cryptochromes	2
Total	21

The development, maturation, and release of the carospores are the next major events in the sexual cycle of *P. catenata*. Phototropins may play a role in the light-dependent regulation of gametogenesis, gamete maturation, and mating competence, directly relating to the genesis and maturation of the carospores. However, the release of carospores requires the

breakdown of the tip of the gametophyte thallus. Photoperiod positively correlated with morphological measures of this portion of the sexual life history, as well: open tips of the gametophyte ($p=0.00742$) and presence of carpospores ($p=0.000127$). Regulation of this process is not specifically known, but evidence exists that it may be mediated by light-sensitive pathways. Cryptochromes, both normal and DASH cryptochromes (*Drosophila*, *Arabidopsis*, *Synechocystis*, and Human), may play a role. If the DASH cryptochromes become down-regulated, a decrease in DNA repair and subsequent increase in DNA damage takes place, which leads to cell death (Chen et al. 2007). The cryptochrome VcCRYp, a cryptochrome identified in the green alga *Volvox carteri*, has also been shown to play a role in cell-type specific regulation of programmed cell death (Danon et al. 2006). This may be the mechanism used to kill the cells at the end of the gametophyte thallus. Both the rhodophyte forms of DASH cryptochromes (from *Cyanidioschyzon merolae*, CRYD) and VcCRYp's (from *Galdieria sulphuraria*, CRY1) are present in the genetic data herein, with CRYD being associated with metallochaperone activity. Metallochaperones are involved with certain reactive oxygen scavenging enzymes, which protect the cells from reactive oxygen species (ROS) (Collén & Davidson 1999). ROS damage DNA, RNA, and proteins and are involved in apoptosis (Simon et al. 2000). Down-regulation of the DASH cryptochrome (CRYD) could be the explanation for the mechanism by which the thalli tips degrade. Two unigenes were associated with DASH cryptochromes and 12 associated with phototropins (Table 6).

The next step in the sexual cycle of *P. catenata* is the germination of the carpospore into the Chantransia. CrPhot has been shown to be directly involved in the regulation of the germination

of *C. reinhardtii* zygotes in response to light (Huang & Beck 2003). In concert with the correlations between many of the growth and life history measurements taken and photoperiod, there is evidence light-induced pathways are involved in the regulation of the sexual stages in *P. catenata*.

Conclusion

The life history of *P. catenata* is well suited for its stream habitat. The perennating Chlantrisia allows it to retain its foothold in the upper reaches of the stream, while the carpospores allow for the continuing recolonization of the stretches of the waterways downstream. From this study, it appears that *P. catenata* uses day length as its primary cue to regulate its sexual cycle, with all measured reproductive stages positively correlating with photoperiod. The light sensitive proteins coded for by PHOT1, CRYD, and CRY1 are implicated in this regulation. They were present in all the functional groups linked to the various stages of the reproductive cycle, and have been linked to these regulatory functions through previous studies.

To further investigate these regulatory cues, studies which collect RNA data at each sampling instance would be valuable. It would allow researchers to elucidate the changes in expression over the growing season, and find a more concrete link between certain genes and the reproductive cycle. This would also allow for RT-PCR to be utilized to investigate the expression levels of specific genes at each sampling instance. Namely genes relating to specific cryptochromes and phototropins, other related genes such as those involved in stress response and ROS scavenging, as well as the unidentified genes which may play some role in the processes being investigated in this study. Also, finding a stream in which *Paralemanea* is

present season after season in the same location would allow a more complete survey of the phenology.

Acknowledgements

I would like to thank the many people who have helped me while undertaking this study. To Dr. Sheath, who allowed me to work in his lab, as well as guided my efforts, exhibited limitless patience while I worked on this project and especially for the pep talks when things seemed hopeless and I was ready to quit with the loss of much of my data. This was my first true laboratory position, and I am grateful for the experience and knowledge that has come from it. Thanks to Dr. Rosalina Stancheva, whose supervision and advice helped me immensely in the lab, and whose drive for precision and organization has made me a better scientist. To Christina Fuller, who helped immensely with various protocols both relating to my lab work and my thesis work, and who offered both professional insight as well as a tremendous amount of personal encouragement throughout my time here.

Without my committee members, this would not have been possible. To Dr. Betsy Read, who helped me refine the genetic portions of this study, guided me with regard to protocols, and generously allowed me use of her lab and equipment. To Dr. William Kristan, who helped me when I had trouble with my statistics and helped me understand the concepts therein.

Thanks also to Dr. Jancovich who took the time to help me understand the concepts of sequence assembly and for being a sounding board for me to bounce ideas off. To Jeani Cressy who helped with equipment, for lifting spirits, and providing welcome breaks in work with conversation. To Dr. Deborah Kristan who helped me traverse the paperwork and requirements

to stay in and finally finish this degree. To my sampling assistant Alex Marr, whose interest in science and enthusiasm helped a great degree.

Finally, to my parents and family who were endlessly supportive of my pursuit of this degree and in every other aspect of my life.

References

Atkinson, G. F. (1890). Monograph of the Lemnaceae of the United States. *Ann Bot*, 4: 177-229, pls 6-9.

Blum, J. L. (1994). *Paralemanea* species (Rhodophyceae) in California. *Proc Indiana Acad Sci*, 103: 1 – 24.

Chan, C. X., Blouin, N. A., Zhuang, Y., Zäuner, S., Prochnik, S.E., Lindquist, E., Lin, S., Benning, C., Lohr, M., Yarish, C., Gantt, E., Grossman, A. R., Lu, S., Müller, K., Stiller, J. W., Brawley, S. H., & Bhattacharya, D. (2012). *Porphyra* (Bangiophyceae) transcriptomes provide insights into red algal development and metabolism. *J Phycol*, 48: 1328 – 1342.

Chaves, I., Pokorny, R., Byrdin, M., Hoang, N., Ritz, T., Brettel, K., Essen, L. O., van der Horst, G. T., Batschauer, A., & Ahmad, M. (2011). The cryptochromes: blue light photoreceptors in plants and animals. *Annu Rev Plant Biol*, 62: 335 – 364.

Chen, J. H., Hales, C. N., & Ozanne, S. E. (2007). DNA damage, cellular senescence and organismal ageing: causal or correlative? *Nucleic Acids Res*, 35: 7417 – 7428.

Choi, S., Hwang, M. S., Im, S., Kim, N., Jeong, W. J., Park, E. J., Gong, Y. G., & Choi, D. W. (2013). Transcriptome sequencing and comparative analysis of the gametophyte thalli of

- Pyropia tenera* under normal and high temperature conditions. *J Appl Phycol*, 25: 1237 – 1246.
- Cole, K. M., & Sheath, R. G. (eds.) (1990). *Biology of the red algae*. Cambridge University Press.
- Collén, J., & Davison, I. R. (1999). Stress tolerance and reactive oxygen metabolism in the intertidal red seaweeds *Mastocarpus stellatus* and *Chondrus crispus*. *Plant Cell Environ*, 22: 1143-1151.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21: 3674–3676.
- Danon, A., Coll, N. S., & Apel, K. (2006). Cryptochrome-1-dependent execution of programmed cell death induced by singlet oxygen in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*, 103: 17036 – 17041.
- Filkin, N. R. & Vis, M. L. (2004). Phenology of *Paralemanea annulata* (Lemaneaceae, Rhodophyta) in an Ohio woodland stream. *Hydrobiologia*, 518: 159-168.
- Flintoft, L. (2011). Transcriptomics: Measuring gene expression in non-model organisms. *Nat Rev Genet*, 12.11: 742 – 742.
- Foster, K. W., & Smyth, R. D. (1980). Light antennas in phototactic algae. *Microbiol Rev*, 44: 572 – 630.
- Furbank, R. T., & Taylor, W. C. (1995). Regulation of photosynthesis in C₃ and C₄ plants: a molecular approach. *Plant Cell*, 7: 797-807.

- Gao, K. S., Ji, Y., & Aruga, Y. (1999). Relationship of CO₂ concentrations to photosynthesis of intertidal macroalgae during immersion. *Hydrobiologia*, 398/399: 355 – 359.
- García-Jiménez, P., & Robaina, R. R. (2015). On reproduction in red algae: further research needed at the molecular level. *Front plant sci*, 6.
- Glick, B. R., & Thompson, J. E. (1993). *Methods in plant molecular biology and biotechnology*. Ann Arbor: CRC Press.
- Gasith, A., & Resh, V. H. (1999). Streams in Mediterranean climate regions: abiotic influences and biotic responses to predictable seasonal events. *Annu Rev Ecol Syst*, 30: 51-81.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B. W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., & Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*, 29: 644–652.
- Graham, L. E. & Wilcox, L. W. (eds.) (2000). *Algae*. Upper Saddle River Press: Prentice-Hall.
- Guiry, M.D. & Guiry, G.M. (2011). *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; searched on 1 November 2011.
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Matther, B. C., Eccles, D., Li, B., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey C. N., Henschel, R., LeDuc R. D.,

- Friedman, N., & Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols*, 8: 1494 – 1512.
- Hambrook, J. A. & Sheath, R. G., (1991). Reproductive ecology of the freshwater red alga *Batrachospermum boryanum* Sirodot in a temperate headwater stream. *Hydrobiologia*, 218: 233 – 246.
- Huang, K. Y., & Beck, C. F. (2003). Phototropin is the blue-light receptor that controls multiple steps in the sexual life cycle of the green alga *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA*, 100: 6269 – 6274.
- Huang, K. Y., Merkle, T., & Beck, C. F. (2002). Isolation and characterization of a *Chlamydomonas* gene that encodes a putative blue-light photoreceptor of the phototropin family. *Physiol Plant*, 115: 613 – 622.
- Iseli, C., Jongeneel, C. V., & Bucher, P. (1999). ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. *Proc Int Conf Intell Syst Mol Biol*, 1999, 7: 138–148.
- Jiménez, J. C., García, M. G. B., & Rodríguez, R. R. (2014). Phenology of *Paralemanea mexicana* (Batrachospermales, Rhodophyta) in a high-altitude stream in central Mexico. *Phycol Res*, 62: 86 – 93.
- Kami, C., Lorrain, S., Hornitschek, P., & Frankhauser, C. (2010). Light-regulated plant growth and development. *Curr Top Dev Biol*, 91: 29 – 66.

- Kanehisa, M., & Goto, S. (2000). *KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res*, 28: 27 – 30.
- Kianianmomeni, A., & Hallmann, A. (2014). Algal photoreceptors: in vivo functions and potential applications. *Planta*, 239: 1 – 26.
- Kugrens, P., & West, J. A. (1973). The ultrastructure of carpospore differentiation in the parasitic red alga *Levringiella gardneri* (Setch.) Kylin. *Phycologia*, 12: 163-173.
- Kunimoto, M., Kito H., Yamamoto, Y., Cheney, D.P., Kaminishi, Y., & Mizukami, Y. (1999). Discrimination of *Porphyra* species based on small subunit ribosomal RNA sequence. *J Appl Phycol*, 11: 203 – 209.
- Lesser, M. P., & Farrell, J. H. (2004). Exposure to solar radiation increases damage to both host tissues and algal symbionts of corals during thermal stress. *Coral Reefs*, 23: 367-377.
- Li, C. Y., Deng, G. M., Yang, J., Viljoen, A., Jin, Y., Kuang, R. B., Zuo, C. W., Lv, Z. C., Yang Q. S., Sheng, O., Wei, Y. R., Hu, C. H., Dong, T., & Yi, G. J. (2012). Transcriptome profiling of resistant and susceptible Cavendish banana roots following inoculation with *Fusarium oxysporum* f. sp. *Cubense* tropical race 4. *BMC Genomics*, 13: 374.
- Mantri, V. A., Thakur, M. C., Kumar, M., Reddy, C. R. K., & Jha, B. (2009). The carpospore culture of industrially important red alga *Gracilaria dura* (Gracilariales, Rhodophyta). *Aquaculture*, 297: 85-90.
- Matusoka, M., Furbank, R. T., Fukayama, H., & Miyao, M. (2001). Molecular engineering of C₄ photosynthesis. *Annu Rev Plant Physiol Plant Mol Biol*, 52:297-314.

- Merchant, S. S., Prochnik, S. E., Vallon, O., Harris, E. H., Karpowicz, S. J., Witman, G. B., ... & Marshall, W. F. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science*, 318(5848), 245-250.
- Onodera, A., Kong, S. G., Doi, M., Shimazaki, K., Christie, J., Mochizuki, N., & Nagatani, A. (2005). Phototropin from *Chlamydomonas reinhardtii* is functional in *Arabidopsis thaliana*. *Plant Cell Physiol*, 46: 367 – 374.
- Prescott, G. W. (1962). *Algae of the Western Great Lakes Area*. Dubuque: WM. C. Brown Company Publishers.
- Reisdorph, N. A. & Small, G. D. (2004). The CPH1 gene of *Chlamydomonas reinhardtii* encodes two forms of cryptochrome whose levels are controlled by light-induced proteolysis. *Plant Physiol*, 134: 1546 – 1554.
- Sage, R., Pearcy, R. W., & Seemann, J. (1987). The nitrogen use efficiency of C₃ and C₄ plants. *Plant Physiol*, 85:355-359.
- Sheath, R. G. (1984). The biology of freshwater red algae. In Round, F. E. & D. J. Chapman (eds.), *Prog Phycol Res*, Vol. 3. Biopress, Bristol, pp. 89-157.
- Sheath, R. G. & Hambrook, J. A. (1990). Freshwater ecology. In Cole, K. M., & Sheath, R. G., (eds.), *Biology of the Red Algae*. New York: Cambridge University Press, pp. 423-453.
- Sheath, R. G., Müller, K. M., Vis, M. L., & Entwisle, T. J. (1996). A re-examination of the morphology, ultrastructure and classification of genera in the Lemnaceae (Batrachospermales, Rhodophyta). *Phycol Res*, 44: 233-246.

- Sheath, R. G., & Sherwood, A. R., (2011). Phylum Rhodophyta (red algae). In John, D. M., Whitton B. A., & Brook, A. J., (eds.), *The Freshwater Algal Flora of the British Isles*, 2nd ed. New York: Cambridge University Press, pp. 123-143.
- Sheath, R. G. & Vis, M. L. (2015). Red algae. In Sheath, R. G., Wehr, J. D., & Kociolek, J. P. (eds.), *Freshwater Algae of North America: Ecology and Classification*, 2nd ed. New York: Academic Press, pp. 237-264.
- Simon, H. U., Haj-Yehia, A., & Levi-Schaffer, F. (2000). Role of reactive oxygen species (ROS) in apoptosis induction. *Apoptosis*, 5: 415-418.
- Thirb, H. H., & Benson-Evans, K. (1984). The effect of light intensities and wavelengths on carpospore germination and the apical tips of the red alga *Lemanea* (Bory 1808). *Nova Hedwigia*, 37: 669 – 682.
- Thirb, H. H. & Benson-Evans, K. (1985). The effect of water temperature, current velocity and suspended solids on the distribution, growth and seasonality of *Lemanea fluviatilis* (Linnaeus) C. Agardh, Rhodophyta, in the River Usk and other South Wales rivers. *Hydrobiologia*, 127: 63-78.
- Vis, M. L., Carlson, T. A., & Sheath, R. G. (1991). Phenology of *Lemanea fucina* (Rhodophyta) in a Rhode Island river, USA. *Hydrobiologia*, 222: 141–146.
- Vis, M. L., & R. G. Sheath, (1992). Systematics of the freshwater red algal family Lemnaceae in North America. *Phycologia*, 31: 164–179.

- Victoria, F. C., Maia, L. C., & Oliveira, A. C. (2011). In silico comparative analysis of SSR markers in plants. *BMC Plant Biol*, 11:15.
- Von Stackelberg, M. V., Rensing, S. A., & Reski, R. (2006). Identification of genetic moss SSR markers and comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biol*, 6:9.
- Wang, B., Guo, G., Wang, C., Lin, Y., Wang, X., Zhao, M., Guo, Y., He, M., Zhang, Y., & Pan, L. (2010). Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res*, 15: 5075–5087.
- Wang, W. (2010). A review on the processing and analysis of next-generation RNA-seq data. *Prog Biochem Biophys*, 5:834-846.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet*, 10: 57-63.
- Wang, Z. Y., Fang, B. P., Chen, J. Y., Zhang, X. J., Luo, Z. X., Huang, L. F., Chen, X. L., & Li, Y. J. (2010). De novo assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweet potato (*Ipomoea batatas*). *BMC Genomics*, 11:726.
- Wehr, J. D., Sheath, R. G., & Kociolek, J. P. (Eds.). (2015). Freshwater algae of North America: ecology and classification. Elsevier.

- Wolfe-Simon, F., Grzebyk, D., & Schofield, O. (2005). The role and evolution of superoxide dimutases in algae. *J Phycol*, 41:453-465.
- Xie, C., Li, B., Xu, Y., Ji, D., & Chen, C. (2013). Characterization of the global transcriptome of *Pyropia haitanensis* (Bangiales, Rhodophyta) and development of cSSR markers. *BMC Genomics*, 14: 107 – 129.
- Yang, H., Mao, Y. X., Kong, F. N., Yang, G. P., Ma, F., & Wang, L. (2011). Profiling the transcriptome of *Porphyra yezoensis* with Solexa sequencing technology. *Chin Sci Bull*, 56:2119-2130.
- Ye, J., Fang, L., Zheng H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., & Wang, J. (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Res*, 34: W293–297.
- Zar, J. H. (1984). *Biostatistical Analysis*, 2nd ed. New Jersey: Prentice Hall.
- Zhang, G., Guo, G., Hu, X., Zhang, Y., Li, Q., Li, R., Zhuang, R., Lu, Z., He, Z., Fang, X., Chen, L., Tian, W., Tao, Y., Kristiansen, K., Zhang, X., Li, S., Yang, H., Wang, J., & Wang, J. (2010). Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res*, 20: 646–654.
- Zhang, J. N., Liang, S., Duan, J. L., Wang, J., Chen, S. L., Cheng, Z. S., Zhang, Q., Liang, X. S. Q., & Li, Y. R. (2012). De novo assembly and characterization of the transcriptome during seed development, and generation of genetic-SSR markers in Peanut (*Arachis hypogaea* L.). *BMC Genomics*, 13:90.