Marshall Plan Scholarship

Report

Differential expression of meiosis genes during different stages of reproduction in *Daphnia pulex*.

Prepared for the Austrian Marshall Plan Foundation

Submitted by: Marelize Snyman

Supervisor UIBK: Dr. Markus Möst, Ph.D.

Dr. Birgit Schlick-Steiner, Ph.D.

Supervisor UTA: Dr. Sen Xu, Ph.D.

Innsbruck, October 2018

Acknowledgement

First of all, I would like to express my gratitude towards the Austrian Marshall Plan Foundation for giving me the opportunity to participate in this amazing program, and to be able to work on my project at the University of Innsbruck. Without the funding it would not have been possible. It was a great experience, I learned a lot and made very valuable connections.

I would also like to thank the Ecology Department at the University of Innsbruck, especially Dr. Birgit Schlick-Steiner, for allowing me to use their facilities so that I can conduct my research. A special thank you to Dr. Markus Möst, my supervisor, for your guidance and advice during my stay, and also for agreeing to be part of my dissertation committee at my home institution. I am looking forward to working together in the near future.

I would also like to thank my home supervisor, Dr. Sen Xu, for allowing me to take the summer to participate in this program and for your valuable input and guidance.

I would also like to express my gratitude to Oswald Jenewein for informing me about this wonderful opportunity, and for your support and guidance. It is it much appreciated. I would also like to thank Sandra Scherl, from the International Office at the University of Innsbruck, for providing me all the information regarding the program, helping me get in contact with the faculty at the Ecology Department, and also for regularly checking up on me to make sure that I was having a good experience. I really appreciate it.

This opportunity certainly has been a highlight and I am very grateful to everybody who helped and supported me. Once again, thank you.

Information

Name: Marelize Snyman

Host University: University of Innsbruck, Austria

Home University: University of Texas at Arlington, USA

Department: Department of Molecular Ecology

Supervisor UIBK: Dr. Markus Möst, Ph.D.

Dr. Birgit Schlick-Steiner, Ph.D.

Supervisor UTA: Dr. Sen Xu, Ph.D.

Abstract

Epigenetics is the study of the effects that a fluctuating environment has on gene expression without an alteration of the genetic code. Very little is currently known on how epigenetics affects important mechanisms including gene expression, development, tissue differentiation and reproduction. The freshwater microcrustacean, *Daphnia pulex*, has been well studied in terms of toxicology, ecology and evolution. What makes *Daphnia* a highly suitable organism for the study of epigenetics is that most of the clones reproduce via cyclic parthenogenesis. All offspring produced parthenogenetically will be genetically identical to one another as well as their mother, indicating that any changes in development or reproduction has to be determined epigenetically. The goal of this study was to determine the transcriptional differences of meiosis or meiosis related genes during various stages of reproduction in order to gain a better understanding of the mechanisms leading from sexual reproduction to parthenogenesis and ultimately obligate asexuality. Three *Daphnia pulex* clones were collected at 4 different reproductive stages: early meiosis and late meiosis, leading to resting egg production, and early parthenogenesis and late parthenogenesis, leading to brood pouch egg production. Resting egg production was induced via crowding and limiting the food supply. RNAseq libraries were prepared and sequenced using Illumina HiSeq 2500 with 150bp paired end reads. Sequences obtained were then quantified with Kallisto and analyzed using the Sleuth tool in Rstudio. Based on our results there are clear transcriptional differences between the late ameiotic and late meiotic stages, but less so between the earlier stages of ameiosis and meiosis. When comparing the transcriptional differences between early meiotic and late meiotic stages there was also a clear difference, however the early ameiotic and late ameiotic stages had little variance. The meiosis or meiosis related genes that contributes the most to the expression differences regulates microtubule binding and movement, coordination of cell division as well as chromosome segregation.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Reproduction: Moving towards asexuality.

In most animal taxa the prevailing mode to produce offspring is sexual reproduction despite the two-fold cost associated with it. Benefits that may contribute to the persistence of sexual reproduction include the ability to purge deleterious mutations from the genome and to increase genetic diversity which can in turn lead to adaptability. Some taxa also have the ability to reproduce asexually while others have even developed a strategy to combine the different modes leading to alternating cycles of sexual and asexual reproduction termed cyclic or facultative parthenogenesis (Simon *et al.* 2003).

The development of asexuality is a question that has long baffled many scientists throughout the years. Currently there are 4 different known origins of parthenogenesis. Parthenogenesis can have a spontaneous origin and thus happen spontaneously through mutations that occur in sexual lineages. These mutations could take place in genes responsible for sexual forms or in genes required for the successful completion of meiosis. If parthenogenesis has a hybrid origin, meiosis is disrupted and a window is created allowing for the selection of cytological processes to rescue egg production (Simon *et al.* 2003). Hybridization is a very common rout to parthenogenesis as most unisexual vertebrates have a hybrid origin. Hybridization is also very diverse in invertebrates and have been demonstrated in snails, crustaceans and many insects including weevils, stick insects and grasshoppers (Simon *et al.* 2003). Parthenogenesis can also have a contagious origin where some obligate parthenogenetic clones retain the ability to produce males which are genetically identical to the females produced along with them. Most of these males, unlike their sisters, retain the ability to produce viable haploid gametes via meiosis and can thus mate with other sexually reproducing *Daphnia* clones (Lynch *et al*. 2018). This will then lead to the production of both sexual and parthenogenetic females via a contagious fashion through the spread of meiosis suppressor genes from the parthenogenetically produced males to the female offspring. This origin has been developed and validated in *Daphnia* and aphids. A lesser known origin of parthenogenesis is by infection via vertically inherited microorganisms. *Wolbachia*, a member of the proteobacteria, is known to induce parthenogenesis in parasitoid wasps such as *Trichogramma* as well as tripes and mites. Sexual reproduction can be restored in females reproducing parthenogenetically due to a *Wolbachia* infection through treatment with antibiotics (Simon *et al.* 2003).

Within these modes different mechanisms can be utilized to give rise to diploid individuals without the need for fertilization. Apomictic parthenogenesis causes the suppression of meiosis leading to the development of offspring by a mitosis-like cell division from an unfertilized egg. The resulting offspring will then be genetically identical to their mother. This mechanism is more commonly found in invertebrates such as rotifers and all major groups of arthropods (Simon *et al.* 2003). Automictic parthenogenesis, in contrast to apomictic parthenogenesis, retains meiosis however diploidy is restored by the duplication or fusion of gametes produced by the mother. This mechanism is more commonly found in stick insects and some weevils (Simon *et al.* 2003). Gynogenesis is another form of parthenogenesis in which the sperm of a bisexual related specie is needed in order to stimulate egg development. The sperm does not contribute to the genetic material of the offspring and thus they will be genetically identical to the mother (Simon *et al.* 2003). Hybridogenesis is a hemiclonal mode of reproduction where the maternal genes are transmitted clonally, and paternal genes are transmitted

sexually to the offspring (Simon *et al.* 2003). Species can either utilize one of these mechanisms or a combination of them when switching to asexuality. For example, unisexual fish species in the genus *Poeciliopsis* are known to use gynogenesis and hybridogenesis to produce parthenogenetic offspring. Another common mode is facultative parthenogenesis, also termed Tychoparthenogenesis, where eggs would spontaneously develop without fertilization (either via apomixis or automixis) between sexual cycles. This mode of parthenogenesis is very common in insects (Simon *et al.* 2003) and also in crustaceans like *Daphnia* (Schurko *et al.*2009), the organism chosen for this project. Previous studies have indicated that parthenogenetic eggs are produced via apomixis, where mature oocytes undergo a nuclear division similar to somatic mitosis (Sabelli 1972). Later studies, however, indicated that *Daphnia* undergo a type of abortive meiosis where the first meiotic division is arrested during early anaphase followed by a second division similar to the second meiotic division. Thus, *Daphnia* do no adopt typical apomixis (Schurko *et al.* 2009).

1.2 Epigenetics

Epigenetics is the study of changes in organisms caused by modifications in gene expression rather than an alteration of the genetic code. Changes in the epigenome can be inherited from one cell to another and in some instances also from parents to offspring (Arimondo *et al.* 2012). The epigenome influences a wide variety of different processes including gene expression, development, tissue differentiation as well as the suppression of transposable elements. While the genome remains relatively static within an individual, the epigenome can be dynamically altered by fluctuating environmental conditions including predation, food availability and seasonal changes (Massicotte *et al.* 2011). Major epigenetic processes that allow the interaction of the genome with the environment include DNA methylation, histone modifications and RNA interference.

Of the processes mentioned above, DNA methylation is one of the most studied epigenetic phenomena. During the process of DNA methylation, methyl groups are added to the 5' position of cytosine in a DNA molecule. This addition of a methyl group will then prevent the gene from being transcribed due to the repression of transcription initiation. It was first believed that only vertebrate organisms and plants have methylated DNA, however, DNA methylation has recently also been discovered in invertebrate organisms including the freshwater microcrustacean *Daphnia magna,* a sister specie of *Daphnia pulex* (Vandegehuchte *et al.* 2009). During histone modification histone proteins can undergo methylation, phosphorylation, acetylation, ubiquitylation and sumoylation. These changes to the histone proteins can affect gene expression either by altering the chromatin structure or by recruiting histone modifiers. Histone modifications can greatly affect biological processes such as transcriptional activation or inactivation, chromosome packaging and DNA damage or repair (Yun *et al*. 2011). With RNA interference, small RNA's interact with mRNA transcripts causing the transcripts to be split into several fragments. This fragmentation of the mRNA transcripts prevents it from being translated into a protein and the gene from being expressed (Ünlü *et al.* 2015).

With whole genome sequencing becoming more affordable and with the development of modern technologies in recent years, our understanding of epigenetic regulation has increased. We now know that epigenetics does not only plays a role in certain phenomena such as imprinting, position effect variegation or X chromosome inactivation, but also in cellular processes such as development, homeostasis and differentiation (Arimondo *et al.* 2012).

1.3 Daphnia as a model organism

The freshwater microcrustacean, *Daphnia*, has previously been well studied in terms of ecology, toxicology and evolution, but their epigenetic processes remain largely unexplored. It has been previously shown that sex determination, production of males and females and distinct morphological phenotypes are all determined epigenetically (Robichaud *et al.* 2011).

What makes *Daphnia* a highly suitable organism for the studying of epigenetic processes is the manner in which it reproduces. Most of the *Daphnia* species follow a cyclic parthenogenetic life cycle where it is capable of switching between parthenogenesis (asexual reproduction) and sexual reproduction depending on environmental conditions. Some *Daphnia* clones have also been identified as obligate parthenogenetic where they only reproduce asexually (De Meester *et al.* 2009). Previous studies have indicated that cyclic populations experiencing favorable conditions, such as a low population density and high food availability, undergo a modified oogenesis where the first meiotic division is aborted and only a second mitosis-like division takes place resulting in the production of diploid eggs which can then develop in the absence of fertilization (Harris *et al.* 2012). Other than the occurrence of rare events such as mitotic recombination, conversion and mutation, all offspring produced from these diploid eggs will be genetically identical to one another as well as their mother (Robichaud *et al.* 2011). The offspring will not necessarily be epigenetically identical to one another or their mother because epigenetic modifications are influenced by the environment that they are exposed to (Robichaud *et al.* 2011). When stressful conditions such as high population density and a lack of food arise some of these clonally produced eggs will develop into males rather than females. Stressful conditions also lead to the establishment of conventional meiosis in cyclic populations where haploid eggs are produced by the females which first need to be fertilized by haploid sperm produced by the males before they are deposited into the ephippium (Xu *et al.* 2014). In obligate parthenogenetic populations favorable conditions will lead to the production of parthenogenetic eggs which will get deposited into the brood pouch to develop as in cyclical populations. During unfavorable conditions males are also produced parthenogenetically along with other females, however they serve no reproductive function. Under stressful conditions the eggs that are deposited into the ephippium are not haploid as in cyclical populations, instead they are diploid and thus fertilization is not needed in order for them to develop when favorable conditions arise again (De Meester *et al.* 2009).

With sex determination and reproduction being determined epigenetically, parthenogenetically and sexually reproducing females are genetically identical, and are also genetically identical to their mother. Parthenogenetically produced males are thus also genetically identical to the parthenogenetically produced females (Robichaud *et al.* 2011).

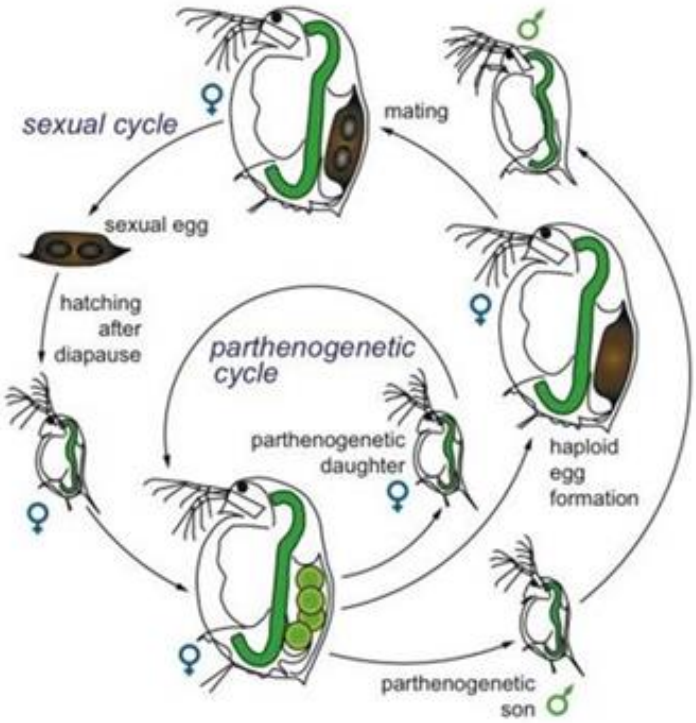


Figure 1: Life cycle of cyclic parthenogenetic Daphnia (Crustacea: Cladocera) showing alternating parthenogenetic and meiotic reproductive cycles (Ebert, 2005). During parthenogenesis all clones produced will be genetically identical to one another as well as the mother. Once unfavorable conditions arise males start to develop as well as haploid eggs and sperm via regular meiosis so that sexual reproduction can occur.

With these changes occurring in populations that are genetically identical and reproduce parthenogenetically, any changes in reproduction or sex determination has to be epigenetically determined (Robichaud *et al.* 2011). This opportunity to study epigenetic effects in the absence of genetic variability makes *Daphnia* the perfect organism to study epigenetic processes and to get a better understanding of how the environment helps to regulate the genome (Robichaud *et al.* 2011).

1.4 Project description

For this project we are interested in determining which meiosis or meiosis related genes are actively being transcribed in obligate parthenogenetic *Daphnia pulex* clones. Specifically, we will be comparing gene expression between the early and late developmental stages of eggs that will get deposited into the brood pouch versus those that will get deposited into an ephippium. When comparing obligate and cyclical parthenogenetic clones, the eggs produced parthenogenetically under favorable conditions are believed to occur via the same mechanism. The resting eggs produced under stressful conditions are hypothesized to occur via a different mechanism since the eggs in obligate clones are diploid, and in cyclic clones are haploid. Resting eggs in cyclic populations thus first needs to be fertilized before they are deposited in the ephippium in order for development to be possible later on. Since the results obtained in this study will at a later stage be compared to the results obtained by performing the same study on cyclical clones, reproduction leading to eggs being deposited into the brood pouch will be termed as

having undergone parthenogenesis or ameiosis. Eggs that are deposited into the ephippium will be termed as having undergone meiosis or sexual reproduction even though in our *Daphnia pulex* clones they are diploid. This will ensure that the correct stages are compared in studies down the line and to avoid confusion since both eggs in obligate populations are produced parthenogenetically. Through this comparison at a later stage we will not only get a better picture of the genes involved in allowing a switch in reproductive modes, but also the genes allowing for the establishment of obligate asexuality.

A list of meiosis or meiosis related genes have been obtained and was complimented by using BLAST to identify other meiosis or meiosis related genes present in *Daphnia* that is commonly found in yeast (*Saccharomyces cerevisiae*). In order to obtain a better understanding of when these genes are actively being transcribed, the different modes of reproduction were divided into 4 different stages: early parthenogenesis, late parthenogenesis, early meiosis and late meiosis. Each of these stages could be determined by viewing the organism under a light microscope and looking at the color and shape of the ovary, and if there is any indication of an ephippium.

During early parthenogenesis a small thin green ovary extending along the gut with thin even oil droplets is visible. During development the ovary will then become bigger and bulkier and during late parthenogenesis, right before the eggs are deposited into the brood pouch, one can start to distinguish some of the individual eggs (Figure 2C). These eggs are diploid in both cyclic and obligate *D. pulex* populations.

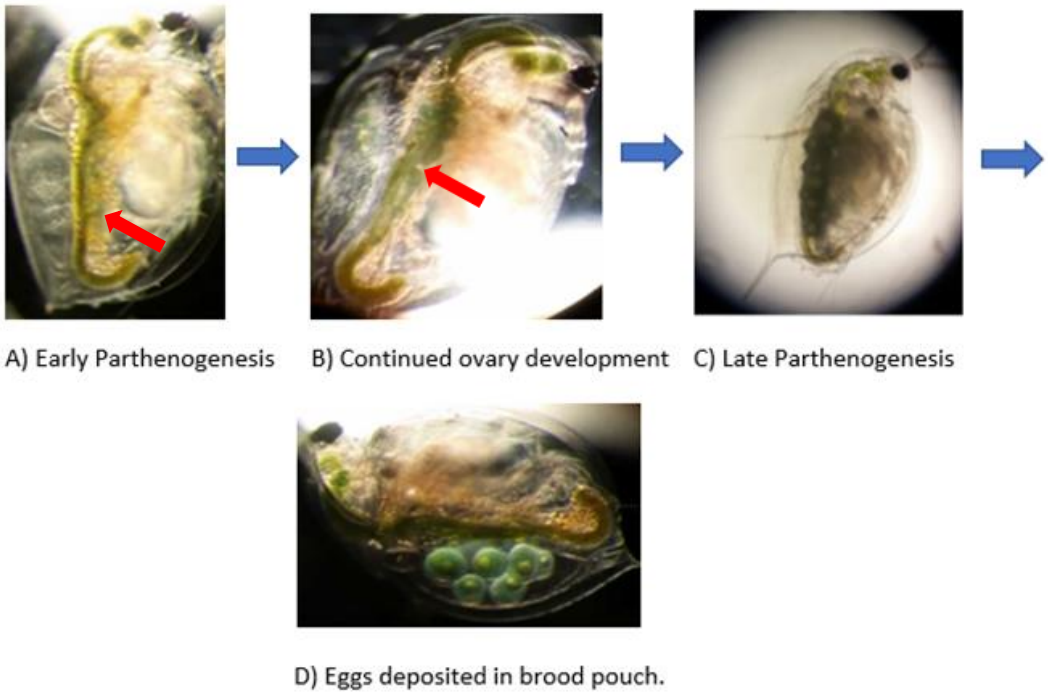


Figure 2: Parthenogenetic ovary development. During early parthenogenesis one can see a small green ovary starting to develop along the gut. During development the ovary will become bigger and bulkier up until late parthenogenesis where one can start to distinguish individual eggs. After late parthenogenesis the eggs are deposited into the brood pouch on the back of the *Daphnia* where they will develop into Daphniids.

During early meiosis the ovary is a milky brown color along the edge of the gut. As the development progresses the milky brown ovary will start to extend along the gut and will also become bigger. During late meiosis one can view a bigger milky brown ovary and also early signs of ephippia development (Figure 3). In cyclic *D. pulex* populations eggs deposited into the ephippium first need to be fertilized in order to be diploid. In obligate parthenogenetic *D.pulex* populations the eggs are diploid and do not require fertilization before being deposited.

Figure 3: Meiosis ovary development. During early meiosis the ovary is a milky brown color along the end of the gut. During development the milky brown color will extend along the gut and with late meiosis one can also see signs of early ephippia development.

Key differences between the two different modes of reproduction is: during parthenogenesis the ovary can be semi-transparent or a green color, one is able to distinguish eggs in the ovary before they are deposited into the brood pouch, and there is no ephippia production. Meiotic ovaries have a milky yellow/brown color and one is not able to distinguish eggs while they are developing in the ovaries. During late meiosis one can also see the early development of the ephippium into which the eggs will be deposited once developed.

After a sufficient number of animals have been collected at each reproductive stage, RNAseq analysis will be used along with our list of meiosis or meiosis related genes to determine which genes are actively being transcribed during which of the four reproductive stages. By comparing the different stages of the obligate parthenogenetic clones, and later comparing them to the results obtained from cyclical populations, we will get a better understanding of the genes involved in allowing the organism to switch from sexual to asexual reproduction. This method will also allow us to determine which genes might be involved in establishing obligate asexuality.

1.5 Data Analysis

For data quantification the program, Kallisto, was used and for analysis, Sleuth, was used in Rstudio.

Kallisto is a recently developed RNAseq quantification program based on the pseudoalignment of reads and fragments. Previously quantification was done by using programs such as TopHat2 to aligning reads to a reference genome or transcriptome, and then utilizing the companion program Cufflinks to estimating transcript abundance. This process however has been proven to be very time consuming and could become restrictive with an increase in samples. For example, aligning 20 samples with 30 million RNAseq reads using the TopHat 2 program takes about 28 core hours on 20 cores, while quantification with Cufflinks can take an additional 14 hours. Even though methods such as streaming algorithms or naïve counting of reads can be used to speed up the quantification of aligned reads, these methods usually lead to a decrease in quantification accuracy (Bray *et al.* 2016).

What makes Kallisto very useful is that it circumvents the alignment step. In stead of aligning transcripts to a reference genome or transcriptome and then calculating abundance, Kallisto pseudoalignes reads to a reference and then provides a list of transcripts that are compatible with each read while avoiding the alignment of individual bases. Thus, aiming to identify the transcripts from which the reads could have potentially originated from. It has been shown in previous studies that accurate quantification does not depend on where inside the transcripts the reads originated, but instead on which transcripts generated them. This has shown to be very useful as previous studies was able to analyze 30 million RNAseq reads in just under 10 minutes using a standard laptop computer (Bray *et al.* 2016).

The careful evaluation of the variability in gene expression from samples to identify relevant expression differences between conditions remains crucial for differential analysis of RNAseq experiments. In order to identify differences in transcript or gene expression, the tool Sleuth was used in Rstudio. Sleuth analyzes differential gene expression by using bootstapping along with response error linear modeling in order to decouple biological variance from inferential variance (Pimentel *et al.* 2017).

2. Methods

2.1 Maintenance of Daphnia

All of our *Daphnia pulex* clones were obtained and kept in the laboratory in continuous parthenogenetic reproduction under the following standardized conditions: 20 ℃ with a 16h:8h light dark cycle. The clones were fed a standardized algae mixture containing *Scenedesmus obliquus* (*Acutodesmus obliquus*) twice a week, and the artificial medium (Daphnia COMBO) in which they were cultivated was changed once a week. The Daphnia COMBO was made by adding 19L of Nanopure water and 20mL of Calcium Chloride, Magnesium Sulfate, Sodium Bicarbonate, Sodium Metasilicate, Boric Acid, Potassium Chloride, Animal Trace Elements and Sodium Selenite respectively (Kilham *et al.* 1998).

2.2 Daphnia collection

First we randomly selected three different *Daphnia pulex* clones to be used in the experiment. Six 1L beakers filled with artificial medium and 10 animals were created per clone. For each of the clones 3 beakers were continued under standardized conditions as described above to act as the control. In the remaining three beakers stress was introduced for each clone by allowing an increase in population density and thus a decrease in food availability while keeping the other factors constant. The animals were then screened twice a day under a light microscope and collected for RNA extraction at the various stages of reproduction: early parthenogenesis, late parthenogenesis, early meiosis and late meiosis. For each reproductive stage 3 tubes filled with 400uL of RNA later and 15-20 animals were collected representing 3 biological replicates and thus creating a total of 36 samples. The samples were stored at -20℃ until the collection was completed.

2.3 mRNA isolation and Quantification

Once the collection was completed each sample was thawed and washed twice with 500uL of cooled RNAse-free water to remove all of the RNA later. The animals were then transferred to 2ml tubes containing 800uL of RNA lysis buffer and ZR Bashing Beads from the RNA isolation kit (Zymo). The samples were then lysed with the use of a TissueLyser II (Qiagen) for 300 s at 30 Hz. RNA isolation was then continued by using the RNA isolation kit (Zymo) and by following the manufacturer's instructions.

Once RNA extraction on all 36 samples were completed, 12 random samples were chosen to be analyzed using the Bioanalyzer (Agilent 2100) and the Eukaryote Total RNA Nano Series assay. This was done to ensure that the extracted RNA was of good quality and did not degrade. Results obtained from the Bioanalyzer in Figure 4 shows the presence of strong bands around the 2000kb region. These bands indicate the presence of the ribosomal subunits 18S and 28S. Ideally one wants to see high-weight peaks for these subunits,because if they are of low molecular weight or flattened, then degradation has occurred which will affect downstream processing.

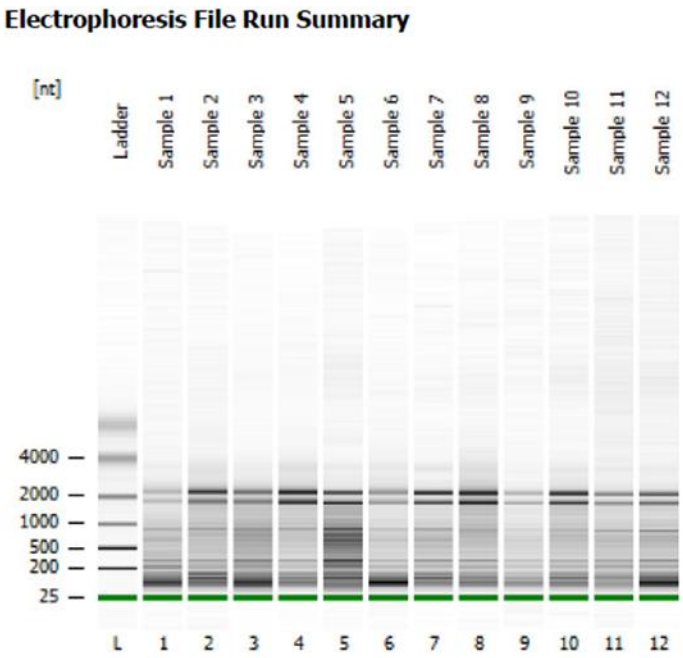


Figure 4: Electrophoresis summary after first extraction obtained by running 12 random samples on the Agilent 2100 Bioanalyzer using the mRNA Nano Series assy.

From the Bioanalyzer results displayed in Figure 4, samples one, six and nine was re-extracted due to too much degradation occurring. Samples two, four, five, seven, eight and nine was considered to have high quality RNA with minimal degradation. After the RNA quality was verified, mRNA was purified from the total RNA by using the NEBNext Poly (A) mRNA Magnetic Isolation Module. The quality of the mRNA was once again verified afterwards using the Bioanalyzer and Qubit.

2.4 cDNA library construction and Sequencing

The cDNA libraries were prepped by using the NEBNext Ultra RNA Library Prep Kit and following the manufacturer's instructions. Index codes were then ligated in order to identify each individual sample. First and second strand cDNA synthesis was performed followed by a purification step of the double-stranded cDNA using 1.8X Agencourt AMPure XP Beads. End repair of the cDNA library was then performed followed by adaptor ligation and another purifying step using AMPure XP beads. DNA fragments ligated with adaptor molecules were then amplified in a 15 cycle PCR followed by another round of purification using Agencourt AMPure XP Beads. The quality of the library was then verified on the Bioanalyzer by using the Agilent High Sensitivity Chip and by performing a round of qPCR. The samples were then sent off for sequencing using Illumina HiSeq 2500 with 150bp pair end reads.

2.5 Data Analysis

The first step when utilizing Kallisto is to create a transcriptome index which is necessary for pseudoalignment. A transcriptome index was created by using a reference transcript file for the *D. pulex* clone PA42. The reference transcript file was created by using information obtained from wfleabase.org. Afterwards the abundance of the transcripts was calculated, the Sleuth tool was used for quantification. All of the commands used during data analysis were obtained from the Patcher Lab website (https://pachterlab.github.io/kallisto/ and https://pachterlab.github.io/sleuth/).

To create the transcriptome index from the reference transcript file, transcripts.fasta.gz:

```
kallisto index -i transcripts.idx transcripts.fasta.gz
```

To quantify the abundance of the transcripts with the two read files (reads_1.fastq.gz and reads_2.fastq.gz) and 100 bootstrap runs:

```
kallisto quant -i transcripts.idx -o output -b 100 reads_1.fastq.gz reads_2.fastq.gz
```

After quantification with Kallisto, the Sleuth tool was used in Rstudio to perform the differential expression analysis. The following steps were used:

```
library(sleuth)
```

#A variable, sample_id, was created to specify where the results were stored:

```
sample_id <- dir("C:/Users/Marelize/Desktop/Data_kallisto")
```

#To display the results, the following command was executed:

```
Sample_id
```

#A list of paths to the kallisto results indexed by sample IDs was collected and combined:

```
kal_dirs <- file.path("Data_kallisto", sample_id)
```

An auxillary table describing the relationship between the Kallisto directories and the samples, as well as the experimental design was loaded:

```
s2c <- read.csv(file.path("RNAseq_table.csv"), header = TRUE, stringsAsFactors=FALSE)

s2c <- dplyr::select(s2c, sample = Sample, Condition)

s2c

s2c <- dplyr::mutate(s2c, path = kal_dirs)

print(s2c)
```

#The Sleuth object was first initialized with the following command:

```
so <- sleuth_prep(s2c, extra_bootstrap_summary = TRUE)
```

Fitting of the full model:

so <- sleuth_fit(so, ~Condition, 'full')

Fitting of the reduced model:

so <- sleuth_fit(so, ~1, 'reduced')

#The test was then performed with the following:

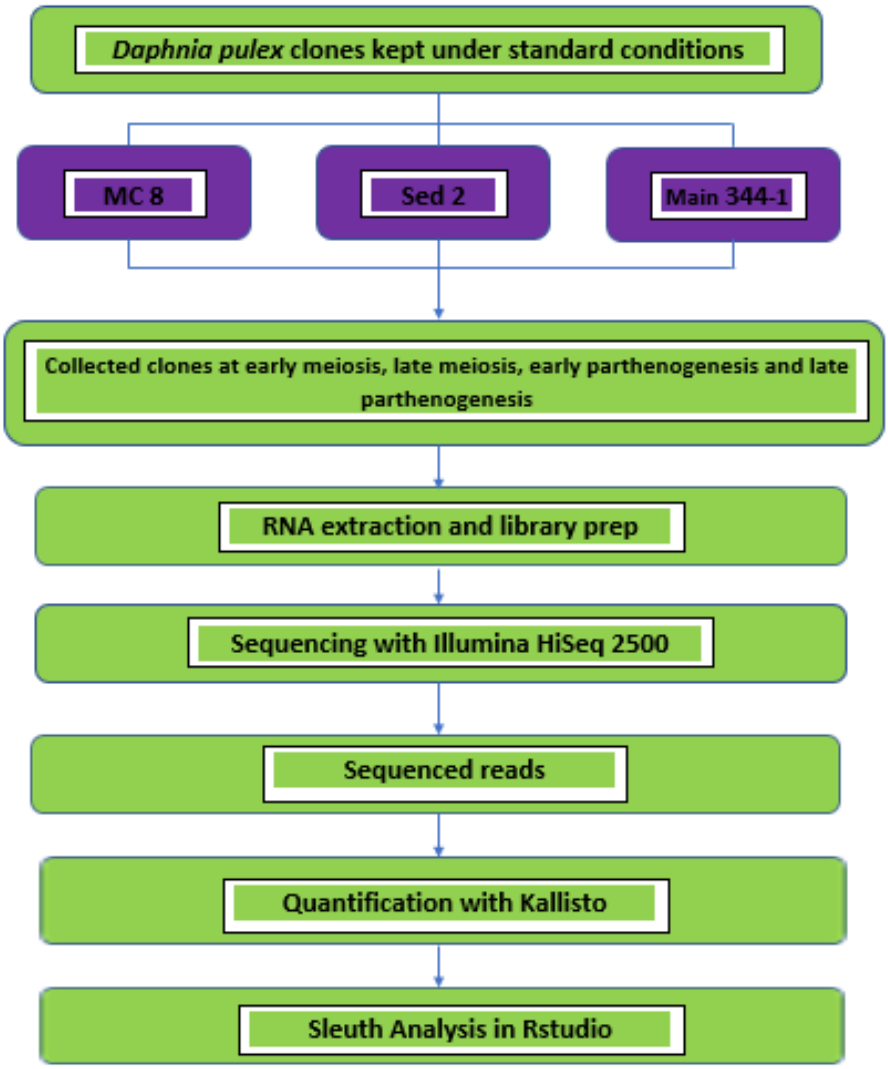
so <- sleuth_lrt(so, 'reduced', 'full')

#The following command was then used to view and interact with the results:

sleuth_live(so)

Sleuth live was used for exploratory analysis and to create the PCA plots which allowed for visualization of the variance between the samples.

2.6 Schematic summary of sample prep, sequencing and analysis



3. Results

After the transcripts were quantified using Kallisto, the Sleuth tool in Rstudio was utilized to examine the patterns of differential expression between the transcripts from each sample. Figure 5 is a PCA plot showing the variance between all of the samples analyzed except for B8, which was determined to be an outlier. From this graph there is a clear grouping of the early meiosis samples represented by the green color. The late meiosis samples, represented by the purple color, are also separated out however they appear to be more scattered. From this graph there doesn't seem to be a distinct difference between the early ameiotic and late ameiotic stages.

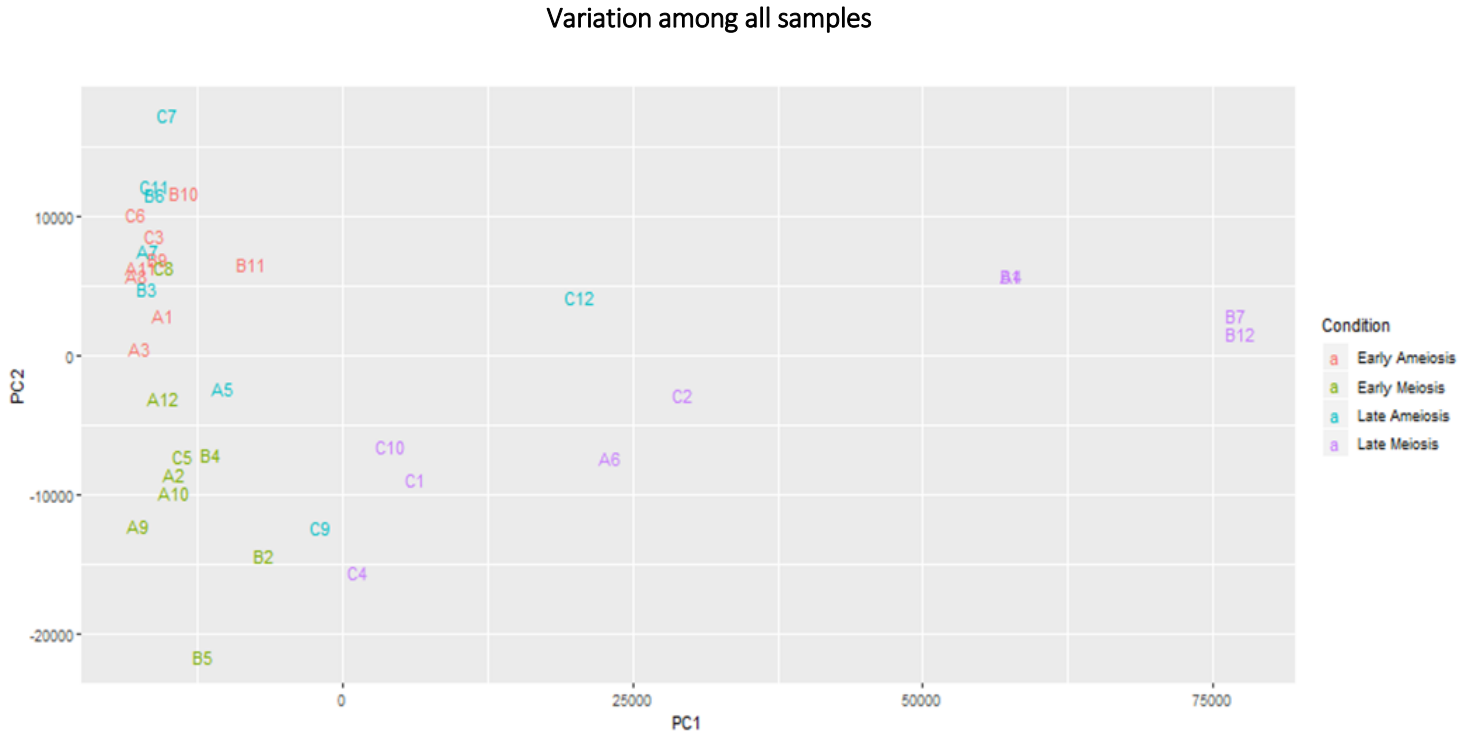


Figure 5: PCA plot showing the variance between each individual sample. The units are in transcripts per million and around 62% of the observed variance can be explained by PC1. In this analyses sample B8 was identified to be an outlier and was thus removed from the analyses.

In the following PCA plot the samples were pooled together according to their condition or reproductive stage. Once again, the sample B8 was excluded from the analysis because it was deemed to be an outlier. In this graph the grouping of the stages can be more clearly visualized.

Variation among pooled samples

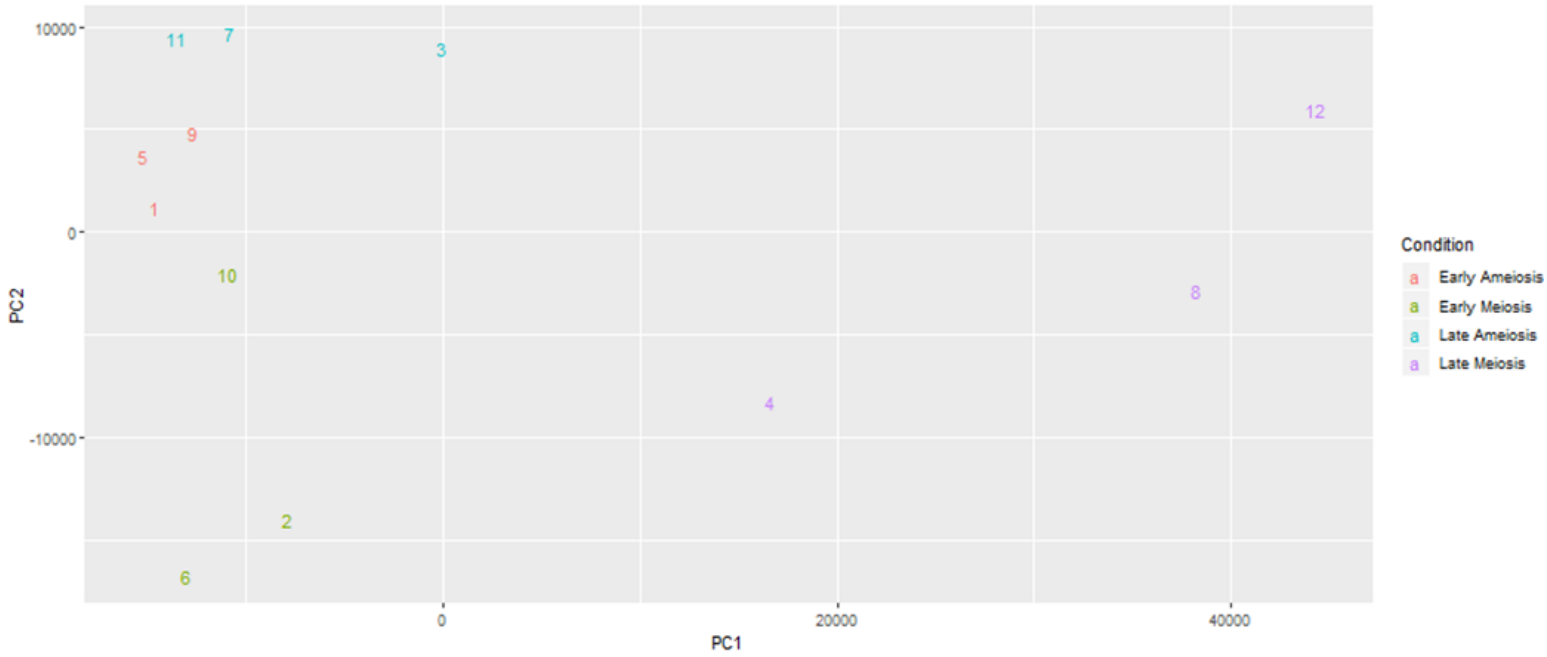


Figure 6: PCA plot in which the 3 different clones were pooled together according to their reproductive stage. The units are in transcripts per million and around 62% of the observed variance can be explained by PC1. In this analyses sample B8 was identified to be an outlier and was thus removed from the dataset.

Table 1 lists the top 10 transcripts which contribute to the variance between all of the samples as explained by PC1. For each transcript the gene from which it is transcribed, along with its function is also listed.

Table 1: The top 10 transcripts contributing to the variance explained by PC1 among all stages

Transcript	Gene	Transcript	Gene
mRNA 16257 (gene 16257)	2-domain hemoglobin protein subunit oxygen binding; heme binding	mRNA 866 (gene 866)	Hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction

mRNA 7229 (gene 7229)	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex	mRNA 4940 (gene 4940)	cop9 signalosome complex subunit 1 isoform x1protein binding
mRNA 7951 (gene 7951)	NA	mRNA 158 (gene 158)	40s ribosomal protein s25 cytosolic small ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis
mRNA 18343 (gene 18343)	Hypothetical protein DAPPUDRAFT_231491 structural constituent of cuticle		
mRNA 3796 (gene 3796)	Hypothetical protein DAPPUDRAFT_300241 (DNA translocase FtsK; Provisional)		
mRNA 7074 (gene 7074)	Cuticle protein, structural constituent of cuticle		
mRNA 9491 (gene 9491)	40s ribosomal protein partial small ribosomal subunit; RNA binding; structural constituent of ribosome; translation; ribosome biogenesis		

The following PCA plot shows the variance between the early meiotic and late meiotic samples. There is a clear difference between the two stages where the early meiotic samples form a clear cluster while the late meiotic samples are more spread out.

Variation between early meiotic and late meiotic samples

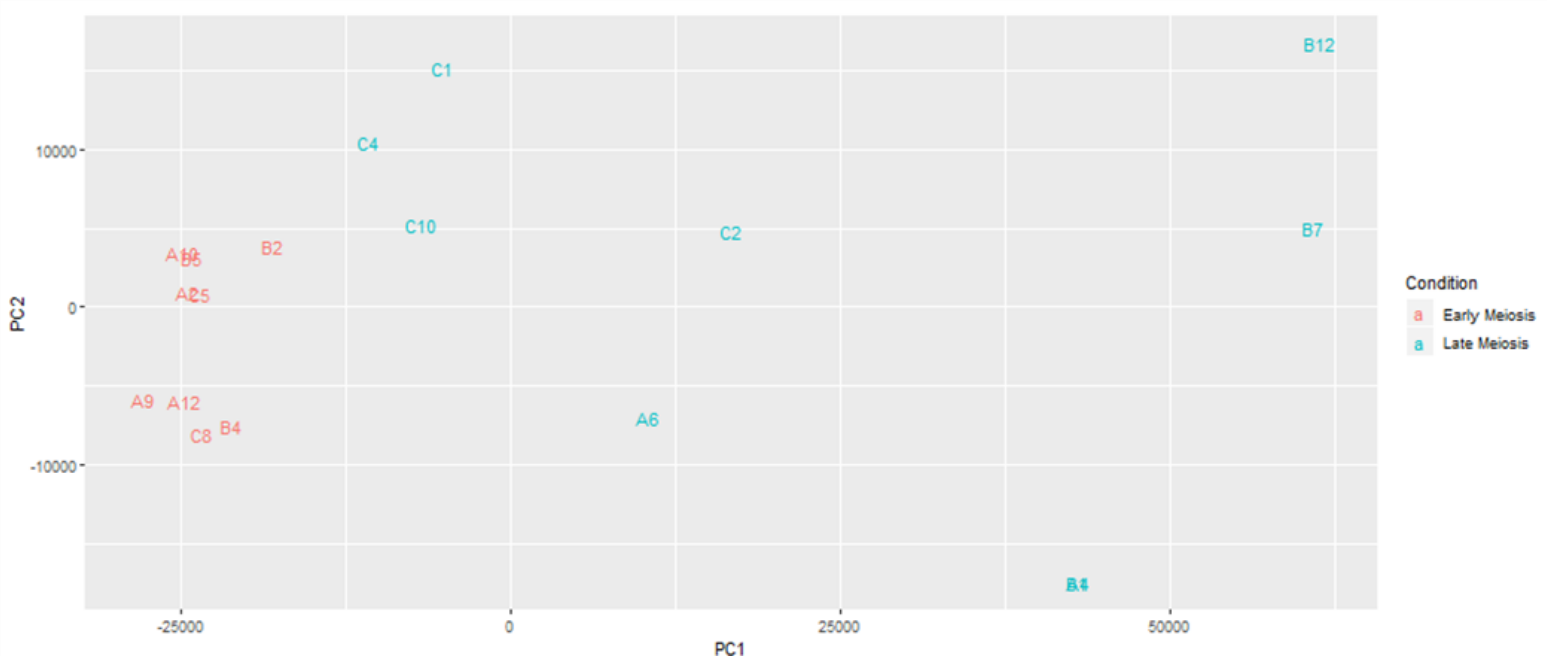


Figure 7: PCA plot showing the variance between the early and late meiosis samples. The units are in transcripts per million and over 70% of the variance is explained by PC1.

Table 2 contains the top 10 transcripts which contribute to the variance as explained by PC1 between the early meiotic and late meiotic stages. The table also contains the gene name from which the transcript was transcribed and also the gene function.

Table 2: The top 10 transcripts that contribute to the variance explained by PC1 between early meiosis and late meiosis.

Transcript	Gene	Transcript	Gene
mRNA18343 (gene 18343)	Hypothetical protein DAPPUDRAFT_231491 (Chitin_bind_4) structural constituent of cuticle	mRNA 3791 (gene 3791)	Hypothetical protein DAPPUDRAFT_300239

mRNA 16257 (gene 16257)	2-domain hemoglobin protein subunit oxygen binding; heme binding	mRNA 9491 (gene 9491)	40s ribosomal protein partial small ribosomal subunit; RNA binding; structural constituent of ribosome; translation; ribosome biogenesis
mRNA 7074 (gene 7074)	cuticle protein structural constituent of cuticle		
mRNA 3796 (gene 3796)	Hypothetical protein DAPPUDRAFT_300241 (DNA translocase FtsK; Provisional)		
mRNA 7229 (gene 7229)	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex		
mRNA 7951 (gene 7951)	NA		
mRNA 866 (gene 866)	Hypothetical protein DAPPUDRAFT_346977 integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction		
mRNA 1838 (gene 1838)	Hypothetical protein DAPPUDRAFT_300242		

The following PCA plot shows the variance among early ameiotic and late ameiotic samples. From the graph there doesn't seem to be a clear grouping or separation between the samples of the two different stages.

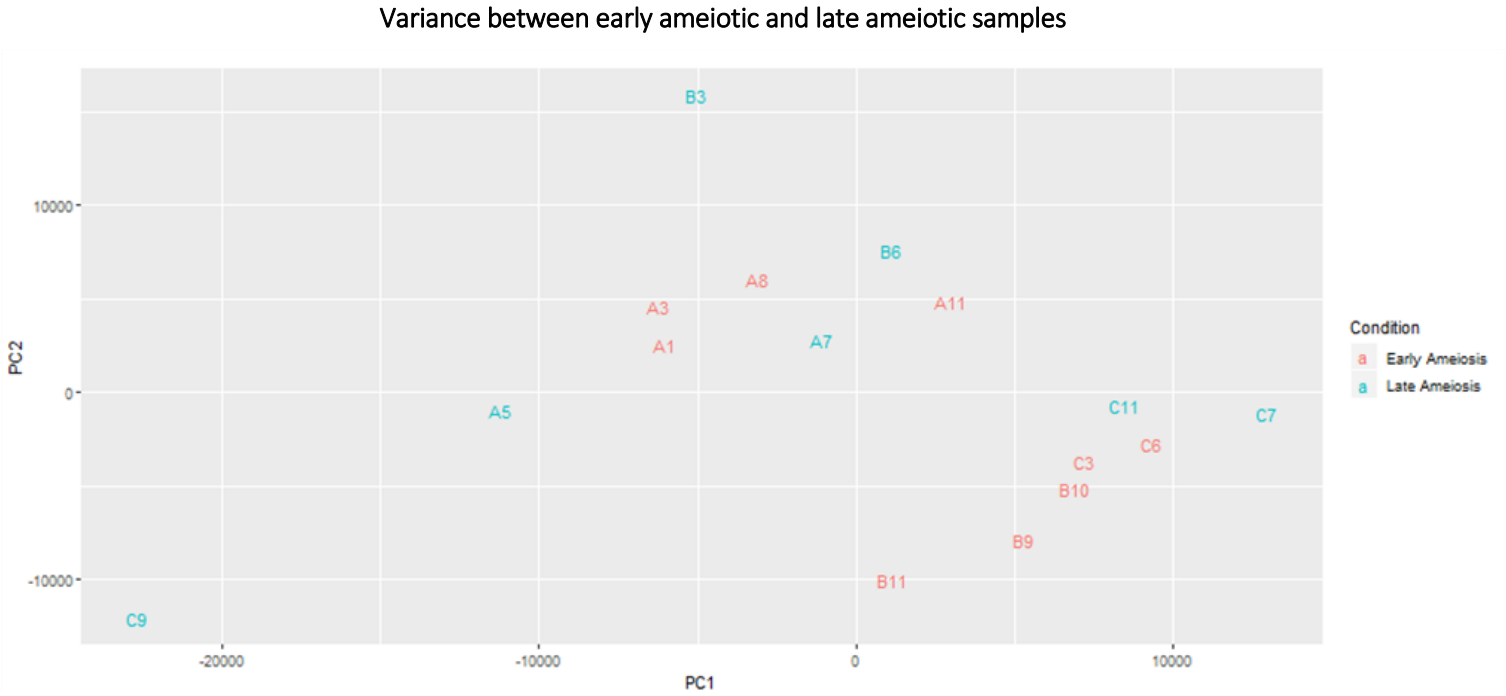


Figure 8: PCA plot showing the variance between the early ameiosis and late ameiosis samples. The units are in transcripts per million and around 25% of the variance is explained by PC1.

Table 3 contains the top 10 transcripts that contribute to the variance between the early ameiotic and late ameiotic samples as explained by PC1. The table also contains the gene name from which the transcript originates as well as its function.

Table 3: The top 10 transcripts contributing to the variance as explained by PC1 between early ameiosis and late ameiosis stages.

Transcript	Gene	Transcript	Gene
mRNA 7229 (gene 7229)	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex	mRNA 17507 (gene 17507)	NA

mRNA 7951 (gene 7951)	NA	mRNA 12889 (gene 12889)	Ribosomal protein l19 ribosome; structural constituent of ribosome; translation; ribosome biogenesis
mRNA 9491 (gene 9491)	40s ribosomal protein partial small ribosomal subunit; RNA binding; structural constituent of ribosome; translation; ribosome biogenesis	mRNA 7224 (gene 7224)	translationally-controlled tumor protein homolog cytoplasm
mRNA 4940 (gene 4940)	60s ribosomal protein l10 ribosome; structural constituent of ribosome; translation; ribosome biogenesis		
mRNA 866 (gene 866)	Hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction		
mRNA 158 (gene 158)	40s ribosomal protein s25cytosolic small ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis		
mRNA 3769 (gene 3769)	60s ribosomal protein l27a large ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis		

The following PCA plot shows the variance between the early ameiotic and early meiotic samples. From the graph on can see that there is a clear division between the samples of the two different stages.

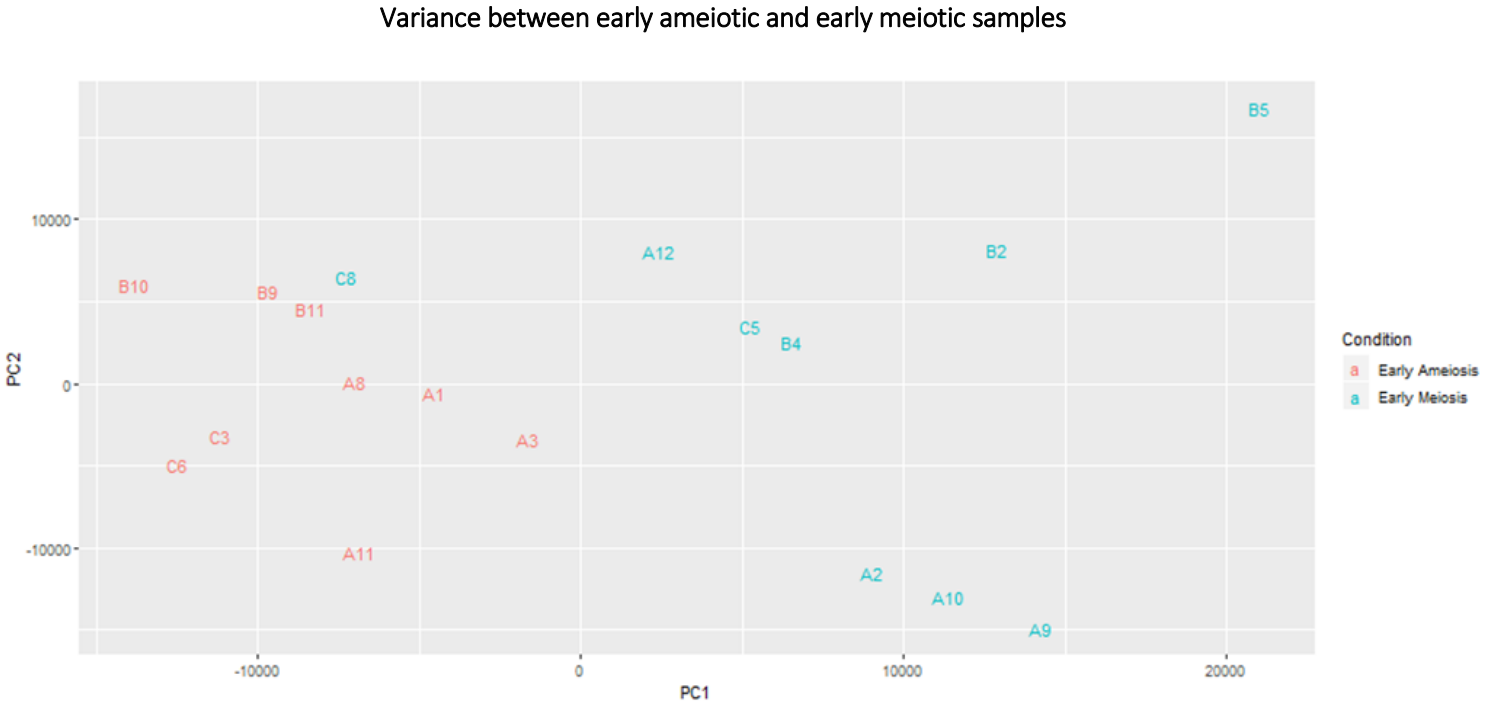


Figure 9: PCA plot showing the variance between the early ameiotic and early meiotic stages. The units are in transcripts per million and around 30% of the variance is explained by the transcripts contributing to PC1.

Table 4 contains a list of transcripts which contribute to the variance between the early ameiotic and early meiotic stages as explained by PC1. The table also contains the gene name from which the transcript originated as well as its function.

Table 4: The top 10 transcripts contributing to the variance as explained by PC1 between the early ameiosis and early meiosis stages.

Transcript	Gene	Transcript	Gene
mRNA 7229 (gene 7229)	Bipolar kinesin krp-130kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex	mRNA 158 (gene 158)	40s ribosomal protein s25cytosolic small ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis

mRNA 7951 (gene 7951)	NA	mRNA 3769 (gene 3769)	60s ribosomal protein l27alarge ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis
mRNA 16257 (gene 16257)	2-domain hemoglobin protein subunit oxygen binding; heme binding	mRNA 12889 (gene 12889)	ribosomal protein l19 ribosome; structural constituent of ribosome; translation; ribosome biogenesis
mRNA 9491 (gene 9491)	40s ribosomal protein partial small ribosomal subunit; RNA binding; structural constituent of ribosome; translation; ribosome biogenesis		
mRNA 4940 (gene 4940)	60s ribosomal protein l10 ribosome; structural constituent of ribosome; translation; ribosome biogenesis		
mRNA 866 (gene 866)	hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction		
mRNA 6951 (gene 6951)	NA		

The following PCA plot shows the variance between the late ameiotic and late meiotic samples. From the graph there is a clear separation between the samples of the two different reproductive stages.

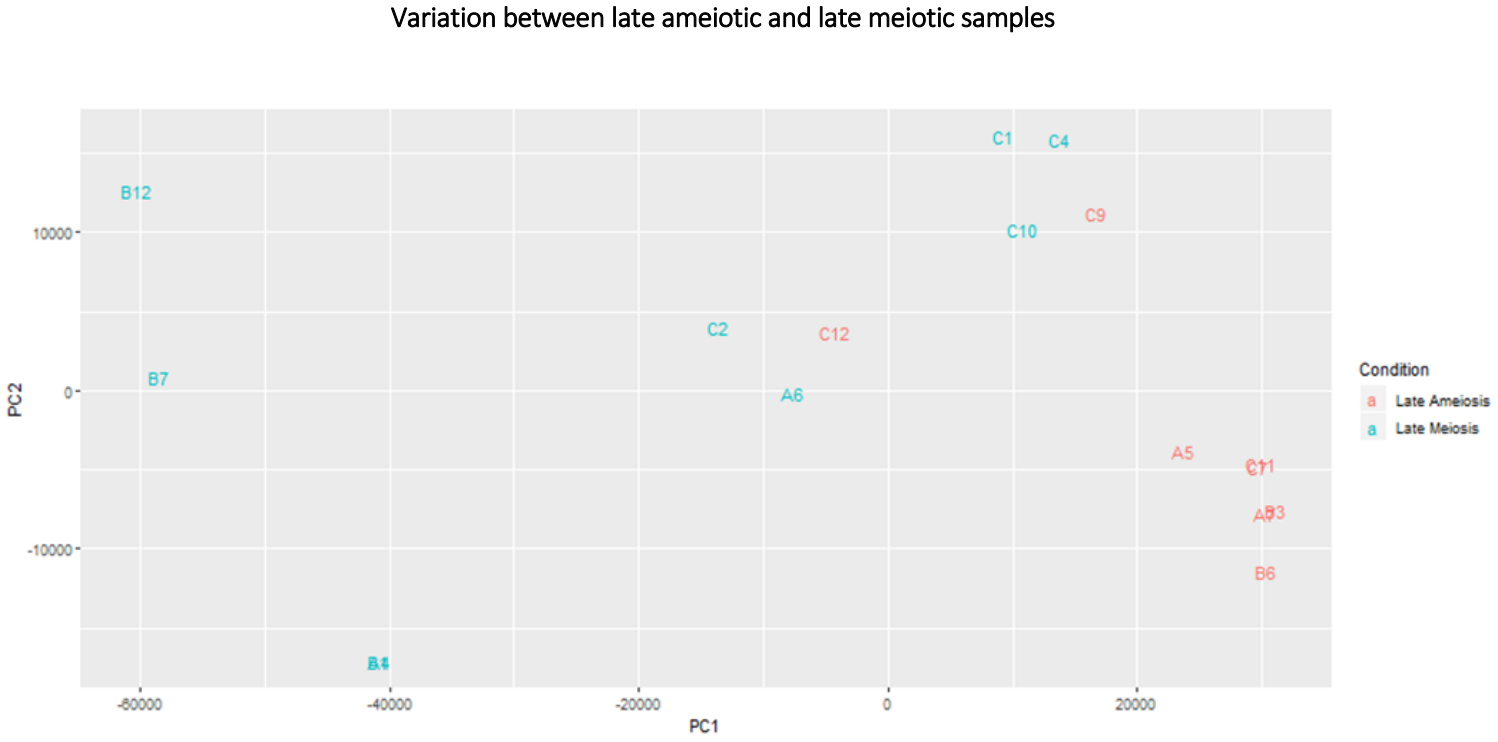


Figure 10: PCA plot showing the variance between the late ameiotic and late meiotic stages. The units are in transcripts per million and around 70% of the variance is explained by transcripts contributing to PC1.

Table 5 lists the top ten transcripts which contribute to the variance between the late ameiotic and late meiotic stages as explained by PC1. The table also has the gene name from which the transcript originates as well as the gene function.

Table 5: Top 10 transcripts contributing to the variance as explained by PC1 between the late ameiosis and late meiosis stages.

Transcript	Gene	Transcript	Gene
mRNA 18343 (gene 18343)	Hypothetical protein DAPPUDRAFT_231491 structural constituent of cuticle (Chitin_bind_4)	mRNA 866 (gene 866)	Hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway;

			visual perception; phototransduction
mRNA 7074 (gene 7074)	Cuticle protein structural constituent of cuticle	mRNA 7951 (gene 7951)	NA
mRNA 3796 (gene 3796)	Hypothetical protein DAPPUDRAFT_300241 (DNA translocase FtsK; Provisional)	mRNA 7075 (gene 7075)	Cuticle protein, structural constituent of cuticle
mRNA 16257 (gene 16257)	2-domain hemoglobin protein subunit oxygen binding; heme binding		
mRNA 7229 (gene 7229)	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex		
mRNA 1838 (gene 1838)	Hypothetical protein DAPPUDRAFT_300242		
mRNA 3791 (gene 3791)	Hypothetical protein DAPPUDRAFT_300239		

Table 6 contains all of the overlapping genes with their gene function when comparing the different analyses. The early and late meiotic stages were compared to the early and late meiotic stages, and the early stag of the two different modes were compared to the late stages of the different modes.

Table 6: Overlapping transcripts between different analysis.

Early and Late Meiotic vs Early and Late Ameiotic		Early Ameiotic and Early Meiotic vs Late Ameiotic and Late Meiotic	
Gene	Function	Gene	Function
Gene 7229	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex	Gene 7229	Bipolar kinesin krp-130 kinesin complex; microtubule motor activity; ATP binding; microtubule binding; microtubule-based movement; microtubule; tubulin complex

Gene 7951	NA	Gene 7951	NA
Gene 9491	40s ribosomal protein partial small ribosomal subunit; RNA binding; structural constituent of ribosome; translation; ribosome biogenesis	Gene 16257	2-domain hemoglobin protein subunit oxygen binding; heme binding
Gene 866	Hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction	Gene 866	Hypothetical protein DAPPUDRAFT_346977 (LOPB9) integral component of membrane; G-protein coupled receptor activity; G-protein coupled receptor signaling pathway; visual perception; phototransduction
Gene 3769	60s ribosomal protein l27alarge ribosomal subunit; structural constituent of ribosome; translation; ribosome biogenesis		

4. Discussion

From figure 5 showing the variance between all of the individual samples there is a clear clustering of the early meiotic samples. The late meiotic samples are also separated from the rest of the samples however they appear to be more scattered. There is a lot of overlap between the early ameiotic and late ameiotic stages indicating that there may not be a lot of variance between the genes being expressed during these two stages. The lack of variation needs to be confirmed by qPCR to ensure that an error did not occur during library preparation or sequencing.

When comparing the early meiotic and late meiotic stages there seem to be a clear separation between the samples as seen in figure 7. From the genes that contribute to the variance as explained by PC1, the following genes have been identified as genes of interest because they play an important role in cell division and chromosome segregation: Gene 3796 and Gene 7229.

Gene 3796 is an Ftsk protein. This protein is an essential cell division protein that aids in the coordination of cell division and chromosome segregation. The N-terminus of the protein aids in the assembly of the cell-division machinery. The C-terminis functions as a DNA motor which moves the dsDNA in the direction of the dif recombination site which is located within the replication terminus region in an ATP dependent manner. The activation of chromosome unlinking by recombination is due to translocation

stops, specifically at the Xer-dif sites, which is where FtsK interacts with the Xer recombinase. The direction of DNA translocation is guided by the FtsK orienting polar sequences (Graham *et al.* 2009).

Gene 7229 plays a role in microtubule binding as well as movement of the microtubule. The differential expression of this gene is interesting because previous research has shown that spindle formation between sexual and asexual organisms differ greatly because of the absence of centrosomes during abortive meiosis (Hiruta and Tochinai, 2012). Little however is still known about how the various systems of spindle formation developed and the role it plays in allowing organisms to reproduce parthenogenetically.

When comparing the early ameiotic and late ameiotic stages as seen in figure 8 there is no clear grouping of the samples and they are not very different from each other. Only around 25% of the variance between the samples are explained by PC1. The transcript that contributed the most to this variance originated from Gene 7229 which plays an important function in microtubule motor activity, microtubule binding and microtubule-based movement.

When comparing the early ameiotic and early meiotic stages as in figure 9, there is a clear divide between the two stages with the early ameiotic samples grouped to the left side of the graph and the early ameiotic samples to the right. Even though there is a clear divide between the two different stages, there is not a lot of variation with only around 30% of the variation explained by PC1. A transcript of interest that contributed the most to this variation originated from Gene 7229 which plays an important role in microtubule motor activity, microtubule binding as well as microtubule-based movement.

When comparing the late ameiotic and late meiotic stages in figure 10 there is a distinct grouping of the late ameiotic stages on the right side of the graph, while the late meiotic stages are on the left and a bit more scattered. Between the samples over 70% of the variance is explained by PC1. Genes of interest that contribute to this variance are Gene 3796, which is an FtsK protein, and Gene 7229 which plays a role in microtubule motor activity, microtubule binding and also microtubule-based movement.

When comparing the genes which overlapped between the early and late meiotic, and the early and late ameiotic stages, Gene 7229 peaked an interest. This gene plays an important role in microtubule binding as well as movement. Since research suggest that *Daphnia* undergo an abortive meiotic process where the first meiotic division is aborted, this gene could play an important role in preventing the separation of chromosomes during the first division, while allowing for a second meiotic division (Schurko *et al.* 2009). When the early meiotic and early ameiotic stages were compared with the late ameiotic and late meiotic stages, Gene 7229 also overlapped. This emphasizes the important role that this gene could potentially play during the abortive meiotic process.

5. Conclusion

Daphnia has the potential to be a great organism for the study of epigenetic processes. They are readily available, very easy to maintain in a lab and their important ecological and evolutionary roles have already been well studied. With *Daphnia* also reproducing via cyclic parthenogenesis one is able to study the epigenetic effects that the environment has on the genome without having to account for underlying genetic variability. Many classical and molecular tools have also been developed to aid in the studying of epigenetics in *Daphnia*.

Despite all of the research that have been done regarding the switching of reproductive cycles, very little is currently known regarding the molecular mechanism and the genes involved. Since most *Daphnia* reproduce via cyclic parthenogenesis and are thus genetically identical, most of the modifications to the genome that allow for this switch has to be determined epigenetically.

For this project we performed RNAseq analysis on 36 samples. The samples were obtained by collecting three different obligate *Daphnia pulex* clones at the four different reproductive stages and having three replicates per stage. Once the sequences were obtained they were quantified via Kallisto and the differential expression was analyzed using the Sleuth tool in Rstudio. Clear differences were observed in gene expression between the early and late meiotic stages. Genes of particular interest that were differentially expressed include gene 3796, a Ftsk protein. This protein is an essential cell division protein that aids in the coordination of cell division and chromosome segregation. The other genes of interest play important roles in microtubule motor activity, microtubule binding and microtubule-based movement. The expression differences between the early and late ameiotic stages were not easily distinguishable and thus qPCR will first be used to verify the results between these two stages.

6. Future Work

For this project we were able to see a distinct difference between the early meiotic and late meiotic stages, in contrast to the early ameiotic and late ameiotic stages where there appeared to be less of a difference. To verify the RNAseq results, Quantitative PCR will be used to amplify the transcripts of single individuals at the different stages. This will especially be a crucial step between the early and late ameiotic stages since the RNAseq results did not show a lot of variance and we would like to confirm that a sequencing error did not occur. Once the results have been confirmed, and a better picture has been obtained of how the genes are differentially expressed, we will move forward with the DNA methylation project.

Since previous studies have not been successful in detecting DNA methylation in the Daphnia specie, *D. pulex*, it was decided to use the bigger sister specie, *D. magna* in which DNA methylation has previously been detected. In 2012, a study done by Harris *et al*. found gene homologues to three main human DNA methyltransferases indicating the occurrence of DNA methylation in *Daphnia magna*. This study will once again allow us to determine which genes are turned on, and thus actively being transcribed during the four different reproductive stages via DNA methylation. During this project, DNA will be extracted from the animals at the same four different stages of reproduction (early meiosis, late meiosis, early ameiosis

and late ameiosis) after which the DNA will undergo bisulfite conversion. After treatment with bisulfite, any unmethylated cytosines will be converted to uracil while the methylated cytosines (5-methylcytosine) will be left intact. Since environmental conditions such as crowding or a lack of food determine which reproductive mode the organism will follow, we suspect that DNA methylation may be one of the mechanisms used to silence genes and allow for the switch from asexual to sexual reproduction.

Once the previous two projects have been completed and a complete list of genes have been obtained, I plan on moving forward with an RNAi project. With this project, RNAi can be used to "knock-out" any meiosis or meiosis related genes to give us a better understanding of what the function of the gene is and the role it plays in allowing the organism to switch from sexual reproduction to asexual reproduction. Small interfering RNA's (siRNA) will be designed to match the list of genes obtained from the previous two projects. The siRNA's will then bind to a special family of proteins known as Argonaute proteins. Once bound, one strand of the dsRNA is removed leaving the remaining strand free to bind to the targeted sequence by following the base pair rules: A binds U, G binds C, and vice versa. The protein can now be used to destroy the targeted messenger RNA by cleaving it. This project will be very helpful in allowing us to determine the function of each gene and the role it plays in reproductive mode switching and the spreading of obligate asexuality.

We are currently only beginning to explore the role that DNA methylation, histone modifications and RNAi play in *Daphnia* development. Further exploration of these three processes and the role that they play in sex determination in this evolutionary and ecologically important organism will not only enhance our understanding of epigenetics, but also have important implications for developmental and evolutionary biology (Harris *et al.* 2012).

Bibliography

Arimondo, P. B., Egger, G., & Tost, J. (2012). epigenetics. Biochimie, 94(11), 2191-2192. doi:10.1016/j.biochi.2012.09.003

Bray, N. L., Pimentel, H., Melsted, P.,Pachter, L. Near-optimal probabilistic RNA-seq quantification, Nature Biotechnology **34**, 525–527 (2016), doi:10.1038/nbt.3519

De Meester, L., & Mergeay, J. (2009). Cyclical parthenogenesis in daphnia: Sexual versus asexual reproduction. (pp. 295-316). Dordrecht: Springer Netherlands. doi:10.1007/978-90-481-2770-2_15

Ebert, D. (2006, October 29). Ecology, Epidemiology, and Evolution of Parasitism in Daphnia. Retrieved from http://www.evolution.unibas.ch/ebert/publications/parasitismdaphnia /ch2f9.htm

Graham, J., Sivanathan, V., Sherratt, D., & Arciszewska, L. (2010). FtsK translocation on DNA stops at XerCD-dif. Nucleic Acids Research, 38(1), 72-81. doi:10.1093/nar/gkp843

Harris, Kami D. M., Nicholas J. Bartlett, and Vett K. Lloyd. "Daphnia as an Emerging Epigenetic Model Organism." Genetics Research International, vol. 2012, 2012, pp. 147892.

Hiruta, C., & Tochinai, S. (2012). Spindle assembly and spatial distribution of γ-tubulin during abortive meiosis and cleavage division in the parthenogenetic water flea daphnia pulex.*Zoological Science, 29*(11), 733-737. doi:10.2108/zsj.29.733

Kilham, S. S., Kreeger, D. A., Lynn, S. G., Goulden, C. E., & Herrera, L. (1998). COMBO: A defined freshwater culture medium for algae and zooplankton. Hydrobiologia, 377(1), 147-159. doi:10.1023/A:1003231628456

Lynch, M., Seyfert, A., Eads, B., & Williams, E. (2008). Localization of the genetic determinants of meiosis suppression in daphnia pulex. Genetics, 180(1), 317-327. doi:10.1534/genetics.107.084657.

Massicotte, Rachel, Emma Whitelaw, and Bernard Angers. "DNA Methylation: A Source of Random Variation in Natural Populations." Epigenetics, vol. 6, no. 4, 2011, pp. 421-427.

Miyong Yun Jun Wu Jerry L Workman Bing Li. (2011). Readers of histone modifications. *细胞研究：英文版, 21*(4), 564-578. doi:10.1038/cr.2011.42

Pimentel, H. J., Bray, N., Puente, S., Melsted, P., Pachter, L. Differential analysis of RNA-Seq incorporating quantification uncertainty, Nature Methods (2017), advanced access http://dx.doi.org/10.1038/nmeth.4324.

Robichaud, N. F., Sassine, J., Beaton, M. J., & Lloyd, V. K. (2012). The epigenetic repertoire of daphnia magna includes modified histones. Genetics Research International, 2012, 174860. doi:10.1155/2012/174860.

Schurko, A., Logsdon, J., & Eads, B. (2009). Meiosis genes in daphnia pulex and the role of parthenogenesis in genome evolution. Bmc Evolutionary Biology, 9(1), 78-78. doi:10.1186/1471-2148-9-78

Simon, J. -., Delmotte, F., Rispe, C., & Crease, T. (2003). Phylogenetic relationships between parthenogens and their sexual relatives: The possible routes to parthenogenesis in animals. Biological Journal of the Linnean Society, 79(1), 151-163. doi:10.1046/j.1095-8312.2003.00175.x

Ünlü, E. S., Gordon, D. M., & Telli, M. (2015). Small RNA sequencing based identification of MiRNAs in daphnia magna. PloS One, 10(9), e0137617. doi:10.1371/journal.pone.0137617.

Vandegehuchte, M. B., Kyndt, T., Vanholme, B., Haegeman, A., Gheysen, G., & Janssen, C. R. (2009). Occurrence of DNA methylation in daphnia magna and influence of multigeneration cd exposure. Environment International, 35(4), 700-706. doi:10.1016/j.envint.2009.01.002

Xu, Shan-Liang, et al. "Identification and Expression Analysis of a doublesex1 Gene in Daphnia Pulex during Different Reproductive Stages." Development Genes and Evolution, vol. 224, no. 3, 2014, pp. 147-157.

Zaffagnini F, Sabelli B (1972). Karyologic observations on the maturation of the summer and winter eggs of Daphnia pulex and Daphnia middendorffiana. Chromosoma 36(2):193–203.