GENE SPLICING IN SQUID MUSCLES

By: Natalie Barco

Under the supervision of Dr. Carol Ely Hepfer and Dr. Jonathan Stoltzfus, two biology students Mireya Lopez Jimenez and Sierra (Baney) Miller have spent multiple semesters researching muscle physiology in squid.

They are focusing on differential expression of genes and muscle isoforms in the squid *Doryteuthis pealeii.* Gene expression is how the information from the gene is used to create a product, which often times are proteins. Their goal is to identify differences that can explain the wide variety of contractile (shortening or tightening movements) properties observed in the obliquely striated muscles of the squid, *Doryteuthis pealeii.*

FIRST, WHAT ARE GENES?

Every cell in an organism contains the same genetic material and thus the same genomic DNA. Not all genes are active, creating a variety of different patterns of gene expression. Different patterns of gene expressions are responsible for the variety of phenotypes, observed characteristics, throughout the organism. The most common way to study gene expression is by comparing expressed sequences to a reference genome, which includes DNA sequences for all genes found in a particular species.

There is no sequenced reference genome for the squid *Doryteuthis pealeii.* Therefore, the research group had to generate a collection of expressed sequences called a transcriptome; they used *de novo* assembly of expressed RNAs. Once a transcriptome is created, it can be used to analyze the relative expression of various RNAs, proteins isoforms, and alternative splicing in squid muscles with distinct contractile properties. These discoveries will advance understanding of how obliquely striated muscle can function in so many distinct ways.

NEXT-GENERATION SEQUENCING

The Research students and their advising faculty had 12 samples of *D. pealeii* that were analyzed using RNA-Seq and over 436 million RNA sequences were generated. RNA-Seq was performed on candidate squid muscles with an Illumina HiSeq sequencer called NGS (Next-Generation Sequencing). This data was used to generate collections of RNA sequences (reads) expressed within each sample.

NGS is a trimmomatic (trinity) software that removes the adaptor and primer from the pair-end reads. Trinity, a *de novo* assembler of RNA-Seq data, was used to generate a reference transcriptome and EdgeR was used to identify differential expression between triplicate

FOOTNOTES

Pair End Reads: Method of DNA sequencing which
gions of DNA which contain repetitive sequences. Pair en
tigs for de novo sequencing by filling in gaps in the given

Isoforms: At least two proteins that have similar functi
acid sequence. They are either encoded by different genes
script from the same gene, but with different exons remov