Effect of the Deletion of 4 HCNEs on *Drosophila melanogaster*

Edward Bull, Class of 2020

Highly conserved non-coding elements (HCNEs) are segments of the genome that have high rates of conservation between taxa that diverged over 4 million years ago. All DNA segments naturally accumulate mutations over time, while sequences that encode important biological functions experience lower rates of mutation due to their functional constraint. Therefore, when sequences that do not even code for proteins have extremely low rates of mutation, this indicates that there must be some functional constraint that causes them to experience such low rates of mutation over millions of years. We tested the hypothesis that these non-coding sequences in some way affect *Drosophila melanogaster* fitness by assaying various elements of flies reared with the deletion of 4 of these HCNEs. The prediction resulting from this hypothesis is that if these conserved sequences do in fact have some kind of biological function, the flies with 4 HCNEs deleted should have significantly reduced measures of fitness and altered gene expression.

The first step of assaying the deletion strain fitness was weeks of collecting virgin flies from both the deletion and intact strains. Flies were collected 3 times a day over the duration of 7 weeks, which were used to assay fitness in the form of female fecundity. 40 virgin females were placed on egg lay plates with 40 males of the same strain, after which eggs were collected over the course of 3 days for 18 replicates. Comparing the deletion to intact flies, the intact on average yielded over 80% more eggs, which supports the hypothesis that these HCNEs play a role in maintaining proper fly fitness.

The second aspect of the project was focused on the genetic level, in which we used qPCR and RNA sequencing to look at the relative expression differences between strains. RNA sequencing identified over 1,000 genes with altered expression levels in the deletion strain, with these genes being distributed across all chromosomes. We also used qPCR to assess the specific amount of gene transcript present for 2 genes- RP49, which is a ribosomal protein present in both samples acting as a control, and a gene with highly differential expression between strains, FBgn0028893, which was identified via RNA seq. RP49 demonstrated nearly identical levels of expression between strains, while FBgn0028893 had a 400-fold difference in expression.

The final project of the summer involved using data available on flybase.org to map differentially expressed genes. First, we manually counted the number of protein coding genes across the entire genome to form a foundation with which to compare the differentially expressed genes. Then, we graphed the distribution of the genes identified via RNA seq. to find regions enriched with affected genes. We found that the genes affected by the deletion were most tightly clumped in the first 2 million and 4 million base pairs of the 2R and X chromosomes respectively. These regions saw 55 to 78 percent of all of their protein coding genes altered by the deletion, which offers an area for possible future inquiry.

Our data collected over the course of the summer seemed to support the hypothesis that HCNEs affect fitness in some way. The fitness assays all offered significant differences in favor of the intact strains, and quantitative analysis of gene transcript levels in flies further indicated that these HCNEs have a wide reaching impact on gene expression across all chromosomes. The looming question that future research could look to answer is the specific function of these sequences. Our data can only indicate if the sequences have some effect on fitness, but not how. Past research has linked HCNEs to enhancer activity and chromosomal mechanics, but these assays would necessitate an entirely different approach. The major result of our research was simply testing the predictions of the hypothesis that these non protein coding sequences play some role in maintaining proper fly fitness, which our data seems to support.

Faculty Mentor: Michael Palopoli
Funded by the INBRE Fellowship