Comparing Putative Enhancer Usage Between Different Species in *Drosophila*

Student Researcher: Justin K. Yang
Advisor: Michael Palopoli

Bowdoin College
Biology Department

A current longstanding debate within the field of evolutionary biology seeks to determine the level of contribution for different drivers of phenotypic variation, either mutations to protein coding sequences or changes to the genome's gene regulation. Cis-regulatory elements known as enhancers have been shown to contribute to phenotypic diversity and lie upstream of the coding region of genes. Enhancers that reside within open chromatin confirmations (as opposed to closed conformation) may actively control the transcription of genes and are known as putative enhancers. Additionally, DNA's conformation has been shown to be heritable, indicating heritability of differing accessibility to enhancers, due to conformational regulation. With this in mind, our study aims to compare putative enhancer usage between different strains of *Drosophilia melanogaster* (fruit flies) as well as other closely related fly species. To do this, we developed a protocol based off of the ATAC-Seq Protocol, first introduced by Buenrostro *et al.* 2013, with the goal of tagmenting DNA from fruit fly brains, that could then be processed into libraries of open chromatin DNA (Buenrostro et al. 2013). Following the sequencing of these libraries by a professional sequencing company, we will then analyze the data using a bioinformatics pipeline generated using the open source SnakePipes pipeline. The work this summer will add to the fly strain datasets that were previously generated by this lab, allowing for a more robust and thorough comparison of the putative enhancer usage between and within fly strains.

The main goal of this summer's project was to continue in-lab work in regenerating DNA libraries from two previously completed natural isolates, as well as generating additional DNA libraries for sequencing. Previously, this lab has generated three DNA libraries of unique *D. melanogaster* natural isolates: Oregon-R (North America), BOG 2 (Bogota, Columbia), Canton-S (Zimbabwe). Two of these libraries (BOG 2 and Canton-S) were found to be of low quality when run through the pipeline, and thus required additional replacement samples to be sequenced to ensure the validity of our data and results. In addition to regenerating DNA libraries for BOG 2 and Canton-S, we also aimed to generate additional libraries from non-*melanogaster* species, notably *D. simulans* and *D. marutiana*. The final goal of this summer was to further our research and understanding about the domains of evolutionary biology and genetics, specifically through research of other studies and previous literature pertaining to ATAC-Seq and our current work.

Our lab used an adapted protocol based off of the Buenrostro *et al.* 2013 ATAC-Seq protocol. The ATAC-seq protocol, created by Buenrostro *et al.* 2013, uses a hyperactive Tn5 mutant transposase to insert large number of a known DNA sequence, known as tags, into sample DNA (Buenrostro et al. 2013; Goryshin and Reznikoff 1998). These tags were matched to primer sequences and used to replicate the sequence of genomic sample DNA containing the transposon. Tn5 is sterically hindered from inserting sequences into DNA bound on the outer shell of histones, and so will enrich genomic sequences within open chromatin conformation (Fig 1.) (Buenrostro et al. 2015). Each natural isolates consisted of 3 samples of DNA libraries. Each sample was generated from the aggregate brain tissue of three to six virgin female flies, that had been aged for two days. Fly brains were dissected by hand in under ten minutes and placed into chilled dPBS buffer. Following dissection, the brains were pestled in lysis buffer then tagmented with the Tn5 transposase. The samples of fragmented DNA were amplified using PCR with uniquely labeled barcodes for each biological replicate and technical replicate (PCR

parameters as described by Buenrostro et al. 2013). Biological samples were divided into two technical replicates each at this step. To reduce the possibility of PCR bias, the samples were run for five cycles, removed, and then held on ice while quantitative PCR (qPCR) was run. Each qPCR technical replicate was run using 10µL of the original samples to determine the optimal number of cycles for each sample. The optimal number of cycles was calculated by running qPCR with the same parameters as PCR for 40 cycles then calculating the number of cycles that corresponded with ¼ of the max fluorescence or point of saturation (Buenrostro et al. 2013). The samples were then returned to PCR following the number of cycles stipulated by qPCR. The amplified DNA was then purified using Qiagen MinElute PCR purification kits (Qiagen), and size selected for 175-1000bp using the Agencourt AMPure XP (Beckman Coulter, Brea, CA) magnetic beads. Samples were then checked for proper size distribution using a bioanalyzer and the corresponding bioanalyzer reagent kits (Agilent Biosystems, Santa Clara, CA). If the samples showed the desired DNA base pair (bp) range as well as bp length distribution, they were quantified using qPCR to determine the concentration of DNA in each sample. After quantification, if all the samples were of the proper concentration for sequencing, they all samples were combined into an equal molar solution and sent for next-generation sequencing (Bauer Core Facility, Cambridge, MA).

This summer, we were able generate an additional two sets of sequencing runs. Each sequencing run consists of thre biological replicate samples, and as such has the possibility of containing one complete set of three biological samples spanning a natural isolate. This summer we produced one sequencing run comprised of two samples of the Canton-S (Zimbabwe) natural isolate and one sample of the BOG 2 (Bogota, Columbia) natural isolates. These samples will be able to be sequenced and then run in the pipeline to replace the previous low quality samples. We were also able to generate a sequencing run of three samples from a *Drosophila mauritiana* natural isolate. As such, this sequencing run is the first set of samples from outside of the *D. melanogaster* species. In addition to the series of samples that were sequenced, two samples of the *D. simulans* 297 strain were also deemed to be of sufficient quality to be sequenced; these samples are on hold for sequencing until another sample is further processed. The sequences were sent off to a sequencing facility and we are still awaiting results.

The generation of these additional samples gives us additional data sets in the comparison of putative enhancer usage between natural isolates and fruit fly species. By regenerating the low quality samples, we increase the accuracy of our data set and ensure that any conclusions drawn have strong and credible evidence to rest upon. The addition of a different species to the data set, should give us a sense of scale for the putative enhancer variation between species, not just within a species. At the moment, the new samples have not been fully sequenced or run through the pipeline with the other samples. As a result, we are not able to draw any conclusions from these data points just yet; however, future work will be done to incorporate and compare these new samples, allowing for a more complete analysis of the putative enhancer variance within and between different fruit fly species.

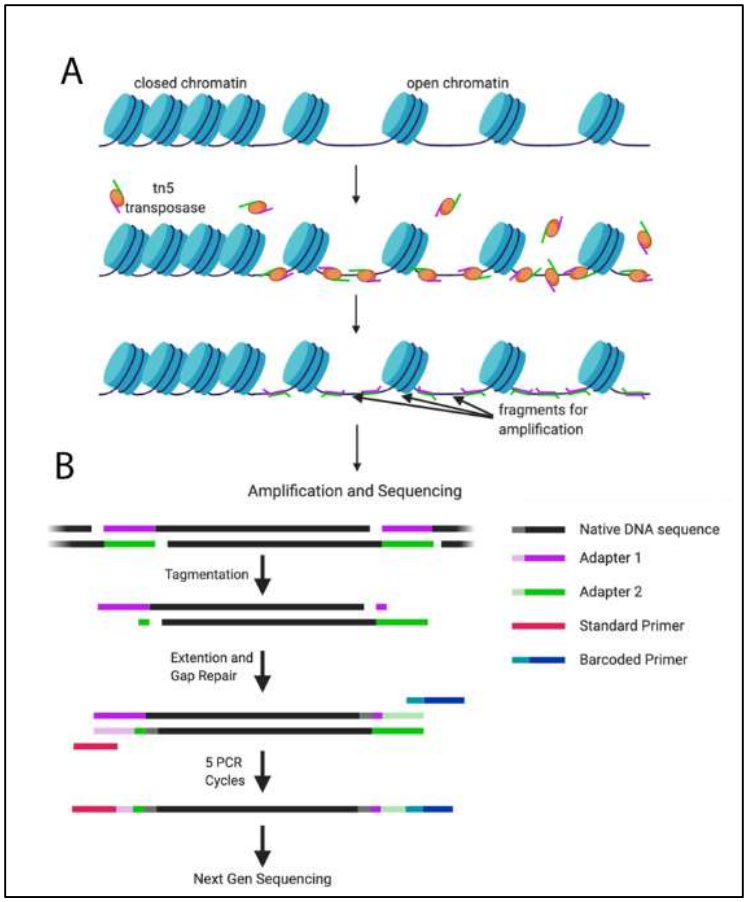

Figure 1: ATAC-seq library preparation. A) Hyperactive Tn5 transposase tagmentation of open chromatin. B) Depiction of the contents of each read before, during, and after PCR amplification. Figure and legend provided from Utku Ferah's Biology Honors Paper "Natural variation in chromatin conformation among populations of *Drosophila melanogaster.*"

I would like to thank our project advisor Professor Mike Palopoli for his insight, guidance, and support. I would also like to acknowledge Mohamed Ashick and LifeBytes, for their work on data analysis and for providing us with the data flow chart. Thank you to DJ Merrill for setting up all the various tools and programs we used on the HPC. Funding for this project was provided by the Maine Space Grant Consortium (MSGC). Bowdoin College is an affiliate of the Maine Space Grant Consortium and any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Aeronautics and Space Administration or of the Maine Space Grant Consortium.

I would also like to thank my friends and previous lab mates: Callie Burkhart, Nicholas Purchas, Grace Pettengill, David Brower, Andy Bolender, and a special thanks to Hannah Konkel for all her help. We would never have been able to continue work on this project if not for all their hard lab work. And thank you to my friends and current lab mates: Laura Yang and Renske Kerkhofs for powering through all the tiring lab days and helping with the last push to get samples for sequencing.

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods. 10(12):1213-1218.

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. Atac-seq: A method for assaying chromatin accessibility genome-wide. Current Protocols in Molecular Biology. 2015:21.29.21-21.29.29.

Goryshin IY, Reznikoff WS. 1998. Tn5 in vitro transposition. Journal of Biological Chemistry. 273(13):7367-7374.