Fitness Effects of Horizontal Gene Transfer from *Wolbachia* to *Drosophila* and Pipeline development and implementation of phylogeny analysis software

Developers: Olga M. Better, Kaitlin Klotz, and John Patterson

**Background/Previous Work:** Trending efforts to understand host-infection relationships to better address pathogenic outbreak has lead researchers into the domain of bioinformatics. Using real life models such as *Drosophila* infection from *Wolbachia, which causes direct phenotype manipulation using horizontal gene transfer,* we can develop a basis by which to analyze the host-infection relationship. *Wolbachia* is a parasite that invades *Drosophila* and causes feminization of the offspring, manipulating cellular processes to favor its own transmission. We expect this phenotype to impact their overall fitness, either positively or negatively.

Previously, researchers in the Rogers Lab at UNCC identified DNA that was transferred from the parasite *Wolbachia* to *Drosophila* fruit flies by isolating sequenced read-pairs that showed one read on the *Wolbachia* chromosome, and one read one the *Drosophila* chromosome. Genuine mutations were supported by multiple pairs of abnormally mapping read-pairs. New genes formed via Horizontal Gene Transfer (HGT) and were detected by Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR), RNA Seq, and fluorescent in situ hybridization (FISH). Careful evaluation of the gene expression patterns of newly transferred *Wolbachia* DNA was implemented to identify cases of new gene formation. A minimum 20 strain subset of the Rogers Lab’s 150 different *Drosophila* strains were used to study why some strains of *Drosophila* thrived after *Wolbachia* HGT and others did not. Phenotypic assays were performed on stressed (temperature shocked, starved, and exposed to chemical mutagens) *Drosophila* to test the effects of *Wolbachia* to *Drosophila* HGT and new gene formation. Finally, the Rogers lab explored the evolutionary dynamics of these novel genetic mutations by examining the phylogeny of HGT segments in *Drosophila*.

**Proposed Improvements/Impacts:** Our proposed improvement to the already-established code is implementing a GUI (Graphical User Interface). To facilitate end user biologists, we propose to develop a more connected and functional software package from the data and code made available by the Rogers group. A GUI interface allows users to interact with electronic devices through graphical icons and visual indicators such as secondary notation, instead of text-based user interfaces. Implementing a GUI and easy file manipulation system should provide a wider user base for this software. This will require an array of software development skills from our group and will require both OO and scripting languages to achieve the desired end product.

**Methods:** Using population genomic sequencing panels for 100 strains of *Drosophila yakuba*, a relative of the model *D. melanogaster* data from the Rogers lab, we want to see if *Drosophila yakuba* can transfer DNA to the nucleus via HGT. We will be searching for mutations that have inserted *Wolbachia* DNA into all 4 major *Drosophila* chromosomes. We will then isolate sequenced read-pairs that have one read the *Wolbachia* chromosome, and one read on the *D. yakuba* chromosome. Next, we will cluster the reads to find the loci where multiple reads match the same locus in the *Wolbachia* genome and the same locus in the *D. yakuba* genome. Genuine mutations will be identified as cases where three or more abnormally mapping read-pairs suggest transfer of DNA from *Wolbachia* to the nucleus. Once mutations are found and scripts are fully functional (i.e. standardized to include more species than *Drosophila*), we will create a GUI to streamline the process for non-programmer users. Included here is a link to the github repository holding the data that will be used in our project: <https://github.com/ryga2016/Wolbachia>. Included in the link is the data that we will use and the code that will be universalized (i.e., reconfigured to be useful with additional species).

**Expected Results:** By the conclusion of the project, we plan to have an organized and user-friendly software package with a GUI for non-programmers that will give us the same results as the original, standalone code produced in the Rogers Lab. Ideally, the standardization of the code from the Rogers lab will result in the ability to apply the package to other species of parasites and hosts and locate novel genes caused by parasitic gene insertions and subsequent mutations.