Kelli McKeegan

Independent Project – White Paper

16 March 2021

**INTRODUCTION TO RESEARCH**

The Junonia coenia densovirus (JcDV) infects multiple species of Lepidoptera, being contracted orally and hindering larval respiration and molting (***1***). For my research, I study variation in the effects of JcDV across populations of a western North American Lycaenid that has only recently been discovered as a permissible host for JcDV: the Melissa blue butterfly (*Lycaeides melissa*). *L. melissa* undergoes 3-4 generations per season (May-September) and is non-migratory, overwintering in their patchy populations as eggs. Not much is known about transmission of JcDV, or even where it is in the wild and in what amount. Although it is nonenveloped, JcDV is a member of the Parvoviridae family, members of which known to be environmentally robust and able to survive in an active state for a period of time outside of the host (***2,3***). However, all host-pathogen relationships are affected by various abiotic (e.g., UV, temperature, wind) and biotic (e.g., tri-trophic interactions, carriers of infection) factors (***4,5***).

In order to begin understanding causes for variation in disease effects as well as routes of JcDV dispersal, one of my projects involves sampling from the surrounding environment in *L. melissa* communities to screen for JcDV with qPCR. I intend to identify the patterns of viral prevalence and load in the context of environmental exposure and ultimately paint the picture of where JcDV is across the landscape of these localities. I predict that *L. melissa* populations with high disease effects will reflect heavy viral presence on the surface of surrounding plants, in the soil, and other arthropods in the community – while populations with low effects might be more naïve to the virus (i.e., have little environmental JcDV exposure). Furthermore, in order to uncover potential temporal cycling, I will compare viral abundance across generations of this multivoltine host. This means that 4 of the 26 sites (ones close to my home base for convenience) was repeatedly sampled throughout the flying season for time series analyses. My prediction for this part is that JcDV abundance in *L. melissa* populations will be low in in the early-season generation (hypothesizing that individuals who “made it” through the winter were resistant to or negative for the virus), and increase with subsequent generations as host populations become more dense and the virus is transmitted.

**DATA & PROBLEM**

Currently this study is between its two stages: Field samples from each site (some being repeated) are collected, but awaiting laboratory qPCR screening. As could likely be assumed, the dataset formed is rather expansive and while I was the only one contributing to it, somewhat messy. There are over 20 columns, inconsistent site labels, a subset of sites that were sampled multiple times, and the data itself has a lot of fluctuation. Due to various climate variables (e.g., summer heat, wind, cloud cover), combined with the multivoltine life history of *L. melissa* (with varying density across generations unique to their habitat), there are some sites and days without data for a certain variable, and some I had to return to because my first attempt was a “bust” or too early. Moreover, a few more columns of data will be added once I run the samples through qPCR: Ct values for two replicates each, along with a column for mean and calculated load from a standard curve regression equation. The problem to be solved lies within the messy data and numerous columns that need to be cleaned and easily accessible for appending and analyzing.

**GOALS**

The goals of my independent project are as follows:

* Read in and clean Excel data sheet
* Easily call numbers or determine how many samples I have of each variable collected
* Add columns once qPCR screening is complete, with math (mean and standard curve)
* Ultimately: Use the tools I have learned in Data Science for Biology to reorganize and manipulate data for convenience of future analyses.

**PROPOSED METHODS**

In order to achieve the aforementioned goals, I will first clean or wrangle the data so that it is consistent, properly formatted, and unnecessary columns or NAs are dealt with appropriately. I will employ a new tool that we will be introduced to soon called OpenRefine for this step. There are a few Python/pandas commands that I might use to summarize database features (e.g., **uniq** to find how many different populations sampled) or to calculate data in columns such as mean and minimum/maximum viral frequency or load. I will then build a relational database from the spreadsheet so that I can easily access specific data for certain variables or sites through queries. We also have yet to learn about relational databases, but it is my understanding this will pre-define relationships between my data and construct the information in a type of layered, table format. This will help for future collections to keep my data organized, as this dataset will be continuously appended in upcoming field seasons whenever I collect for my research. Also, this will help efficiently and easily another table with qPCR data for analysis once I complete that step. Finally, all code and important outputs will be posted to my GitHub repository – which is a whole new achievement of convenience in itself.

REFERENCES

1. Mutuel D, Ravallec M, Chabi B, Multeau C, Salmon JM, Fournier P, Ogliastro M. 2010. Pathogenesis of *Junonia coenia* densovirus in *Spodoptera frugiperda*: A route of infection that leads to hypoxia. *Virology*. 403(2):137-144.
2. Bergoin M and Tijssen P. 2010. Densoviruses: a highly diverse group of arthropod parvoviruses. *Insect Virology.* p. 57-90.
3. Williams, T. 2018. Viruses. In: Hajek N, Shapiro-Ilan D, editors. *Ecology of Invertebrate Diseases*. Hoboken, NJ: John Wiley & Sons. p. 213-285.
4. Ment D, Shikano I, Glazer I. 2018. Abiotic Factors. In: Hajek N, Shapiro-Ilan D, editors. *Ecology of Invertebrate Diseases*. Hoboken, NJ: John Wiley & Sons. p.143-169.
5. Cory JS, Deschodt PS. 2018. The Biotic Environment. In: Hajek N, Shapiro-Ilan D, editors. *Ecology of Invertebrate Diseases*. Hoboken, NJ: John Wiley & Sons. p. 187-212.