**Main Objectives** I aim to conduct field surveys to a) elucidate the distribution, prevalence, and community structure of *Ranavirus* in native amphibian communities of Vermont and b) determine the effects and virulence of the virus on Vermont amphibians; I also aim to c) use a combination of epidemiological modeling and experiments to determine how ranavirus is being transmitted between individuals and between sites and d) use genomic analysis to establish if host and/or pathogen characteristics influence pathogen transmission.

**Question I: What is the distribution, prevalence, and community structure of ranaviruses in amphibian communities of Vermont?**

**Objective I: a) I aim to evaluate the roles that temporal, spatial, environmental, and anthropogenic factors play in *Ranavirus* distribution and prevalence in natural amphibian populations of Vermont, and b) I aim to determine the diversity and abundance of *Ranavirus* species in the state of Vermont.**

**Hypotheses and Predictions**: a) I hypothesize that ranaviruses are present in the state and that prevalence is affected by anthropogenic and environmental factors. I predictthere will be increased prevalence in sites closer to roads and agricultural areas and sites that have higher human visitation and fishing activity. I also predict that ponds with variable water and soil conditions throughout the season will have increased prevalence. b) I hypothesize that viral diversity mimics the viral diversity of surrounding New England states, and predict Frog-Virus 3 will be the primary species, with potentially multiple sub-strains.

**Background**

Surrounding Vermont, ranaviruses and associated mortality events have been found in amphibian communities of Maine (Gahl and Calhoun 2010), Massachusetts and New Hampshire (Green et al. 2002), New York (Brunner et al. 2011), New Jersey (Monson-Collar et al. 2013), and Quebec (Paetow et al. 2011) and Ontario (Greer et al. 2005), Canada. The specific viral species was not reported for most studies, except in New York, which after genetic analysis found FV3-like ranaviruses present, and Ontario and Quebec, where sequences were a 98 and 100 percent match to FV3, respectively. Additionally, prevalence has not been reported for most studies. Quebec found ranavirus prevalence at 85 percent (Paetow et al. 2011), but another study found in 2012 that 100 percent of 18 ponds across 10 states on the United States eastern coast had ranavirus-infected frogs (Crespi et al. 2015).

Although a 1968 New York study found ranavirus in the liver of Northern Leopard Frogs (*Lithobates pipiens*) purchased from a business in Vermont (Clark et al. 1968), no ranavirus surveys have been conducted in the natural amphibian communities of the state. Therefore, ranavirus distribution, prevalence, and its associated factors are virtually unknown. More research is needed to understand why ranaviruses emerge in particular areas (Gray and Chinchar 2015), so the fact that it has not been detected here, as of yet, makes it particularly interesting to study.

**Approach**

*Sampling Design*

In the summer of 2016, I aimed to collect amphibian tissue samples across sites in northwestern Vermont (Figure 1). Sampling would occur at each site, once every other week, from mid-May to August in order to increase the probability of witnessing a mass mortality event and because the sensitivity of PCR using non-lethal tissue samples peaks around 12 days post-virus exposure (Greer and Collins 2007). The potential sites were chosen with James Andrews, the state herpetologist of Vermont. Sites with predicted high amphibian abundance, estimated from Mr. Andrews’ prior survey experience, ideal amphibian conditions (i.e. shallow water and emergent vegetation), assessed either from prior knowledge or from Google Maps satellite view, and those that could reasonably be visited once every other week were selected as contenders. The consistently high amphibian abundance throughout the summer at these sites was considered because an increase in the amount of samples collected would increase the probability of disease detection. A total of 18 sites met the requirements, were chosen as final sampling locations, and were visited a total of seven sampling periods throughout the summer.

*Sampling Techniques*

Previous estimates indicate that for 95% confidence in disease detection and at 5% disease prevalence, tissue should be collected from a sample size of at least 20 individuals (Gray et al. 2015). Therefore, samples were taken from a maximum of 30 individuals at a particular site or after an allotted amount of time, to standardize sampling effort. To obtain a total amphibian search time of an hour and a half ‘person hours’, either two people searched for 45 minutes or three people searched for 30 minutes. All non-endangered species of amphibians at various life stages were collected and placed in either individual plastic bags of distilled water, if a salamander or larval frog, or individual plastic containers, if an adult frog, to limit cross-contamination. Additionally, different pairs of powder-free nitrile gloves were worn while handling each individual. Although lethal methods of sample collection (i.e. liver samples) detect disease more precisely, being able to accurately determine the distribution and prevalence of the disease in Vermont required too large of sample sizes to warrant lethal sampling. Therefore, tail tissue was collected from salamanders and larval frogs, and toe tissue was obtained from adult frogs. Tail collection involved pressing the flat side of a ruler onto the tip of the tail, triggering the natural predator defense of tail autonomy, which minimizes blood loss. Using surgical scissors, one toe was collected from an adult frog per sampling week. If the individual was recaptured, non-adjacent toes were clipped. Forceps and scissors were disinfected between uses with 10% bleach, and waders, boots, and nets were disinfected between sites using 3% bleach, according to the Northeast Partners in Amphibian and Reptile Conservation disinfection protocol (NEPARC 2014). Tissue was stored in 1.5 mL tubes of 90% ethanol in a -20° freezer. A total of 1,822 tissue samples from a total of 10 amphibian species were collected throughout the summer.

*Testing for virus*

DNA was extracted using the Omega Bio-Tek E.Z.N.A. Tissue DNA kit and protocol. I tested for presence of the virus using real-time quantitative PCR (qPCR) because studies have shown that qPCR methods are more sensitive to virus detection when compared to PCR, especially when pathogen concentration is low (Monson-Collar et al. 2013). Real-time quantitative PCR was performed in MicroAmp optical 96-well reaction plates using StepOnePlus Real-Time PCR Systems (Applied Biosystems). The 10uL reactions consisted of 2 uL of 10ng DNA, 5 uL 1X Sso Advanced SYBR Green Supermix (Bio-Rad Laboratories, Hercules, CA, USA), 0.75 uL of each primer, and 1.5 uL of nuclease-free water. Primers RVMCPKim3\_F (5’-TAA- CACGGCATACCTGGACG-3’) and (5’-GATGAGATCGCTG- GTGTTGC-3’) RVMCPKim3\_R (Kimble et al*.* 2014) were used to amplify a 97 bp region of the major capsid protein. The thermal profile consisted of 2 minutes at 95°C, then 35 cycles of 95°C for 20 seconds followed by 65.1°C for 20 seconds. Each individual sample was run in duplicate with four serially diluted standards from 10 to 106 viral copies constructed from cultured Frog Virus-3 (FV3), and a negative control using sterile H2O. qPCR standards were created from cultured FV3 from an outbreak in wild adult Northern Leopard Frogs (*Lithobates clamitans*) from Illinois and obtained from Dr. James Julian at the University of Pennsylvania Altoona. Based on the standard curve, a reaction was considered positive when Ct < 32.0.

*Summer 2017*

If ranavirus is present in Vermont and does not show significant variation in prevalence between weeks, I plan on expanding the number of sites I collect samples from, but reducing my sampling periods to three or four times during the summer. The decrease in sampling periods is due to the fact that mass mortality events were not witnessed during the previous summer. Therefore, reduced sampling may still capture representative prevalence throughout the summer. I plan to record environmental data, such as water and soil quality, from sites each time I collect samples. Additionally, I would like to obtain environmental DNA, in which to test for the presence of ranavirus, in order to determine where else the virus is present besides within the individual.

*Data Analysis: Prevalence*

To address my first hypothesis, virus prevalence can be estimated by dividing the number of cases by the population size. In my case, I will estimate prevalence at a given site and time by dividing the number of infected individuals by the total sample size. I will also be able to compare prevalence estimates between species. Difference in prevalence between sites: chi-squared test (p217; <http://fwf.ag.utk.edu/mgray/RanavirusBook/Chap8/SampleCode_8.5.html>); The change in prevalence between summers: regular ANOVA (ANCOVA?)? Change in prevalence throughout the summer at a site: Repeated measures ANOVA? (changes in mean scores over three or more time points): <https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php>; Logistic regression can also be used to estimate the risk factors associated with ranavirus occurrence among populations (Ranavirus Book p218; <http://fwf.ag.utk.edu/mgray/RanavirusBook/Chap8/SampleCode_8.5.html>)

*Data Analysis: Predicted distribution and potential predictor variables*

In order to attempt to understand what is driving the presence of the virus in certain locations, I can first determine if the means of bioclimatic variables and elevation are different between ponds where ranavirus occurred at some point throughout the summer and those where it did not, using simple one-way analyses of variance (ANOVA). Next, I can use either these coarse BioClim data layers or finer resolution Vermont-specific data layers to create a predicted distribution of the presence of ranavirus in the state. I can start with all 21 predictor layers, determine which are strongly influencing the model, and reduce the number of variables. Since the number of predictor variables may still be more than the number of sites, I can compare the full model to a model that uses a raster PCA of the data layers (how could I interpret this??). Using AIC model evaluation, I can determine the best predictive model and identify potential disease ‘hotspots,’ or areas with a high probability of presence. Next, if a relatively decent model can be constructed for the small area of Vermont, I can conduct a meta-analysis using latitude and longitude points from other studies to form a model for a larger area. Careful selection of other coordinates will need to be made, in order to avoid variance between years and potentially months. Using the driving predictor variables, I can also create a regression tree to see exactly the values that are driving separation between present and absent sites. Additionally, because I plan on collecting environmental information at each site and time point this summer, I will be able to compare water and soil quality values to prevalence numbers. With these types of data, a linear regression will be the best method to determine if there is a relationship between those specific environmental conditions and disease prevalence.

*Data Analysis: Virus diversity and invasion history*

The amplified ranavirus-positive DNA will also be Sanger sequenced at the Advanced Genome Technologies Core Facilities at the University of Vermont. With the sequenced amplified region of the major capsid protein, I can address my second hypothesis and determine which (and potentially how many) species of the six official viruses are present in the Vermont amphibians I collected. In the future, I would like to use next generation sequencing to sequence the whole genome of positive samples and determine which sub-strains are present, as well. Using that collection of ranavirus sub-strains, and potentially sequences along state borders, I can construct a virus phylogenetic tree. From there, I can use a neighbor-joining tree analysis and haplotype distribution to explore the invasion history of the virus into the state. These particular methods have not been used before in the ranavirus system (fly, gen. diversity: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0036176>; spider, human infl.: <http://www.sciencedirect.com/science/article/pii/S1055790303003968>) and has the potential to allow us to better understand how, or if, it is spanning large areas and the best method of limiting its spread. [Mention spatial analysis and phylogenetic study of ranavirus in UK (Price et al. 2016)? They used “twinstim, a function in the R package Surveillance v. 1.7 [38 – 40], to analyse the UK spread of ranavirus-consistent mortality events”]

**Expected Results and Implications**

I expect to obtain two main results from this first question. Firstly, ranavirus has not been documented in natural communities of Vermont, and this is the first large survey of the state. I will be able to present the first instance of the virus, its predicted distribution and potential disease ‘hotspots’ in Vermont, as well as its prevalence and diversity in the sites I sampled. Furthermore, to the best of my knowledge, the predicted distribution of the virus has not been published. For my second main result, I can use a combination of spatial and phylogenetic analyses to present an invasion history of the disease, first locally in the state and then broader-scaled. This method has not been used in this disease system, but has the potential indicate if expansion is human-mediated.