April 2007

A Population Genetics Study of Local Orconectes virilis

Luke N. Adams
Worcester Polytechnic Institute

Matthew P. Basile
Worcester Polytechnic Institute

A STUDY OF THE LOCAL INVASIVE CRAYFISH SPECIES (*ORCONECTES VIRILIS*) USING POPULATION GENETICS.

A Major Qualifying Project

Submitted to the Faculty of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements of the

Degree of Bachelor of Science

by

_____

Luke Adams

_____

Matthew Basile

April 26, 2007

Approved:

_____   _____

Professor Lauren Mathews, Advisor   Professor Destin Heilman, Advisor

_____

Professor Michael Buckholt, Advisor

ABSTRACT

The crayfish species, *Orconectes virilis*, is assumed to be an invasive species in the Massachusetts water systems. This project hypothesizes that the experiments conducted will support the assumption that the species is invasive, by finding low genetic diversity, which may also show evidence of a founder effect. This project investigated 3 new sites, Institute Pond and the French River in Massachusetts, and the Blackstone Gorge in Rhode Island. These sites were analyzed in addition to two sites from a project completed last year by Gottardi and Anderson (2006). The sites were chosen based on their geographic proximity to each other with possibility of crayfish dispersal. This project examines the mitochondrial DNA of the 16S rRNA gene. All the data from the collected crayfish were compared to an outgroup of *Orconectes limosus*. The results show average genetic diversity and a severe case of founder effect, and introduce a new hypothesis that the crayfish we collected are two species, that both resemble, phenotypically, *Orconectes virilis*, and no conclusive assumptions can be made on whether or not the species is invasive.

INTRODUCTION

Obtaining information on the molecular level about ecology, behavior, and most importantly evolution, is the growing field of study called molecular ecology. Molecular ecology employs the techniques of molecular genetics to observe and determine genetic relationships between groups of organisms. These groups can now be reinforced by their genetic content in addition to their physiological and behavioral attributes. Molecular ecology, molecular genetics, and biochemistry will be heavily relied on to conduct this project.

Incorporating and understanding the molecular basis of genes and the way they act has become an important part of the field of molecular ecology (Purugganan and Gibson, 2003). One way to begin to understand this relationship is through the implementation of biochemical techniques. This project uses both an understanding of biochemistry and molecular ecology to determine evolutionary relationships of an invasive crayfish species, *Orconectes virilis*. Genetic links allow scientists to determine ancestral relationships and evolutionary history (Wu, 2001). These links are discovered using molecular genetics and allow more confidence in assigning the organisms into groups as opposed to basing relationships on purely phenotypic information. Placing organisms into groups of other organisms like themselves allows them to be distinguished from other groups, and allows quicker recognition of their phylogenetic information.

The following introduction focuses on further explanation and details about the invasive crayfish *Orconectes virilis*, and about phylogenetics and phylogeography. These core aspects are essential for complete understanding of the goals of this project, which were to understand the relationships of *Orconectes virilis* populations from different

locations in Massachusetts and Rhode Island, and to understand what these relationships suggest about the "invasion" of *Orconectes virilis*.

The crayfish, *Orconectes virilis*, is assumed to be an invasive species in Massachusetts. However, the natural range of this species reaches as far east in the United States as New York, therefore it is possible that Massachusetts may be included in its natural range as well. Given this information we hypothesize that we will find low genetic diversity in the three new populations we investigate. This low diversity may also show evidence of a founder effect, therefore further supporting the assumption that the species is indeed invasive of the water systems in Massachusetts.

Crayfish: *Orconectes virilis*

Crayfish are decapod crustaceans that fall into two families: Cambaridae and Astacidae. The crayfish is also known as a crawdad or crawfish depending on the geographic region. Crayfish is the typical name in the Northeast United States, crawfish in the south, and crawdad in the west. Crayfish have resided on most continents in both lentic and lotic freshwater for millions of years. This habitat provides them with natural protection from predators, for example under rocks, or in small tunnels. They have not had any extreme morphologically changes but their physiology has been upgraded over time. The most beneficial change has been the dispensing of the larval stages and the attachment of the hatchlings to the mother (Holdich, 2002).

Crayfish grow by molting, a process where crayfish shed their exoskeleton. This is because their exoskeletons harden and there is not sufficient room for growth. The age of the crayfish has an effect on how often it molts. As a crayfish ages, the period between its moltings increases from days to months (Crocker and Barr, 1968). After a molting, a

4

crayfish is susceptible to predator attacks, because its new exoskeleton is very soft. During this time, a crayfish will try to hide. Also, crayfish will often eat their molts as a source of calcium (Gilbertson, 1999). Crayfish use molting as a source of limb regeneration as well (Stockard, 1908). Their limbs regenerate because each new exoskeleton has a slightly larger new limb for the crayfish to grow into until it is repaired.

In general, crayfish are made up of two main body sections, the cephalothorax and the abdomen. The cephalothorax is made up of two fused parts, the head and thorax, which has a hard covering called the carapace. Crayfish have five pairs of legs called pereopods. The first pair is modified from the other four and is also known as the cheliped or claw. The other four pairs of pereopods, which are used as walking legs. The abdomen, the second main part, can be folded under the thorax. The abdomen has 5 sets of limbs called pleopods. Their role is to provide movement of water across their gills, which are located in the branchial chambers at both sides of the thorax, so they can receive oxygen. In males the first two pairs of pleopods are modified to aid in sperm transfer. Also, at the caudal end of the abdomen is a telson, which forms a flipper-like tail with the uropods, used for quick maneuvering (Crocker and Barr, 1968).

Crayfish disperse both naturally and artificially. Their movement artificially is primarily from the result of being used as bait, food, or pets. This sometimes allows them to establish themselves as an invasive species. Crayfish will also occasionally disperse on their own, although most prefer to stay in their original habitat. These dispersions were not found to be dependent on size or sex of the crayfish (Byron et al., 2001). The crayfish prefer to move during the summer rather than the winter. Crayfish were found to have a further dispersal rate downstream over upstream (Bubb et al., 2004).

The reproduction of crayfish is found to be seasonal. The reason for this is because of the water temperature. Mating usually occurs during autumn. The eggs are carried through the winter and hatch during spring/summer time (Reynolds, 2002). The eggs are attached to the female's pleopods. The eggs hatch directly into juveniles. In contrast, most marine decapod crustaceans are hatched and released into the open water as planktonic larvae and will undergo significant metamorphoses until they reach the juvenile stage. Juvenile crayfish will cling to their mother for protection and care for a varied period of time before moving on. During this time, the daily movements of an *Orconectes virilis* mother will approach zero (Hazlett et al., 1979). Also in *O. virilis,* when the juveniles separate from the mother, they disperse a distance from one another and from the burrow of the mother (Holdich, 2002).

Orconectes is one the most common genera found in North America. It is made up of 11 subgernera, 81 species, and 11 subspecies (Holdich, 2002). They are mostly used as a fish bait industry and biological supply houses (Huner, 1994). The crayfish that we tested with were of the species *O. virilis.* The species *O. virilis* has a native range from Montana and Utah to Arkansas, New York, and Maine. However, it is found in Massachusetts as an invasive species (Hobbs, 1989). *O. virilis* often have a blue tint to their chelipeds, with the male's chelipeds being larger than the females. They can also be identified by their brown and reddish brown carapace.

Figure 1: Dorsal view of *O. virilis*

Phylogenetics

Phylogenetics is a branch of study that groups organisms according to naturally inherited relationships and depicts natural relationships through the act of grouping these organisms (Donoghue and Kadereit, 1992). The study of phylogenetics is important in determining the connections of families, genera, and species and placing them into a phylogenetic tree (Zimmerman 1934). These phylogenetic trees depict the relationships between a species and other species like it, all leading back to a common ancestral origin. Traditional methods of organism classification have been based on the morphological, developmental, and behavioral characteristics of the organism. However, these traits cannot be used as the only means of classification because they do not always correlate with natural and evolutionary relationships (Lane et al, 1985). The study of

phylogenetics has been around since the time of "On The Origin of Species" by Charles Darwin (1859). The traditional methods of classifying organisms by their phylogenetic relationships were employed for many years, until the recent availability and easy implementation of DNA sequencing into experimental procedures.

The field of phylogenetic analysis is evolving as quickly as the technology used for its study, and currently genetic information is used to determine organism grouping and phylogenetic tree construction. These methods have proved very effective and reliable, which is important for the study of crayfish relationships in this project.

Phylogenetic tree: visual aid for interpreting phylogenetic data

Phylogenetic trees are essential for evolutionary biological studies (Letunic and Bork, 2006). Life forms share common descent with one another. More closely related organisms have more recent common ancestors than the organisms they are distantly related to. Phylogenetic studies are used to discover and reconstruct the correct genealogical ties between organisms and to determine the estimated time of divergence between organisms since their last shared common ancestor (Li, 1997). In phylogenetic studies, the evolutionary relationships among a group of organisms are illustrated by means of a phylogenetic tree (Li , 1996).

A phylogenetic tree is a connected graph composed of nodes and branches, without cycles (Li, 1996). This means that the organisms diverge in one direction, linearly, out from their last common ancestor. The nodes represent taxonomic units and the way the tree branches defines the relationships of the units in terms of descent and ancestry (Li, 1996). The branching pattern is also referred to as the topology and the

taxonomic units denoted by the nodes can be species, populations, individuals, or genes (Li, 1996). The length of the branches is proportional to either the length of time or the number of mutations occurring along the branch (Li, 1996). Figure 1 displays an example of a basic phylogenetic tree, which depicts how each organism diverges from the same ancestral node in no particular scale of time or number of mutations.
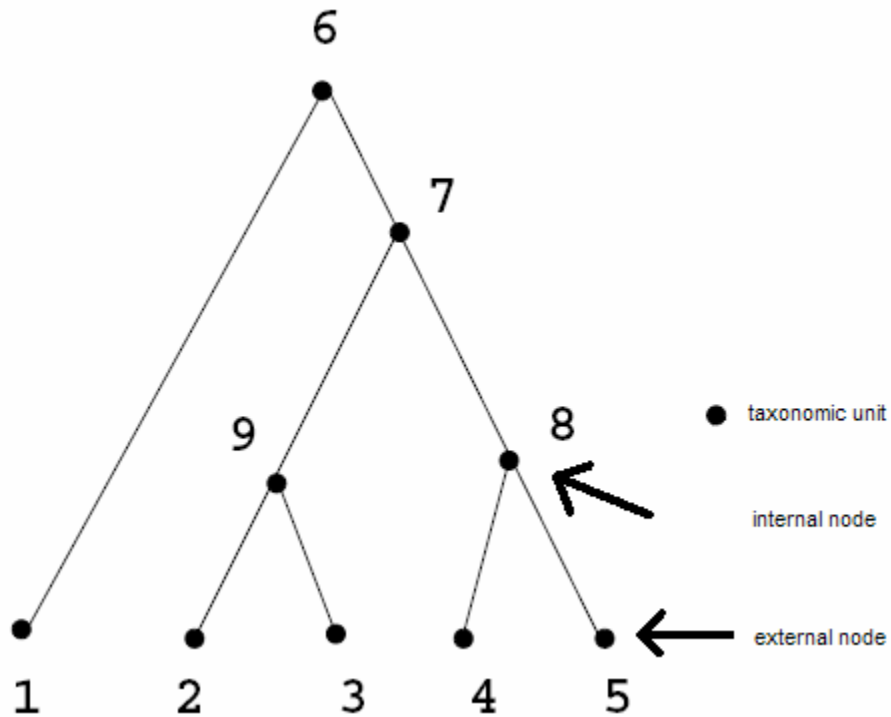

Figure 2: A basic phylogenetic tree (Li, 1996).

Also in Figure 2, there are two types of nodes in the tree. These two types of nodes are the nodes that do not exist at the terminus of the tree branches, and the terminal nodes. The terminal nodes are defined as external nodes and they represent the taxa under consideration, and the non-terminal nodes are defined as internal nodes, which represent ancestors of external nodes (Li, 1996). It is also important to note that evolution occurs independently on the branches after the event of the node (Li, 1996).

9

Phylogenetic trees are either rooted or unrooted. This is determined by the type of information contained in the tree in combination with what the tree is being used to depict by the observer. Rooted trees have a root node with only one path from the root node to all of the external nodes (Li, 1996). Unrooted trees specify relationships among the taxa (Li, 1996). The biggest difference between the two is that the rooted trees, at the root node, have an evolutionary common ancestor and the path of evolution from that ancestor is shown. An unrooted tree, however, does not show the path of evolution, but instead shows relationships between the taxa (Li, 1996). Figure 3 below shows a rooted tree on the left with the root node, and an unrooted tree on the right.
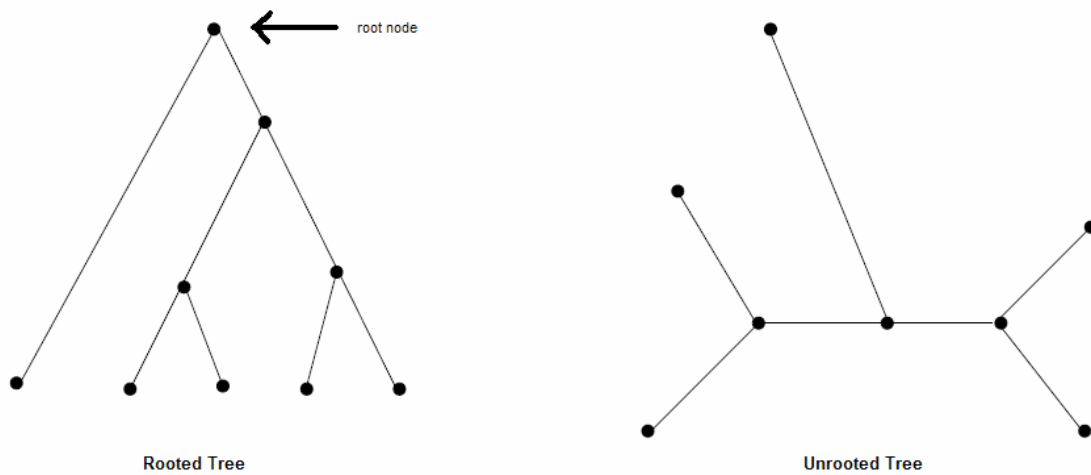

Figure 3: The rooted tree shows the root node and each taxa evolving from it, the unrooted tree shows how closely related the taxa are with respect to each other (Li, 1996).

Furthermore, aside from rooted or unrooted trees, there are three common classes of phylogenetic trees: phylograms, cladograms, and phenograms. The major differences among these types of trees are the length of the branches, the style of the tree, and the order of the taxon depicted. Phylograms summarize the historical or evolutionary

relationships between entities with definite branch lengths which are proportional to the amount of evolutionary change inferred between successive nodes on the tree (Clewley 1998). Cladograms also show evolutionary relationships. However, the branch lengths are uninformative and the order of the branches is significant (Clewley 1998). Finally, phenograms show taxonomic relationships between entities based on their overall similarities without complete evolutionary pathway reconstruction (Clewley 1998). Figure 4 below gives examples of each of these classes of the phylogenetic tree. Only the grouping of the organisms in the phenogram is significant, whereas the order of the organisms in the cladogram is important, but the branch length is not.
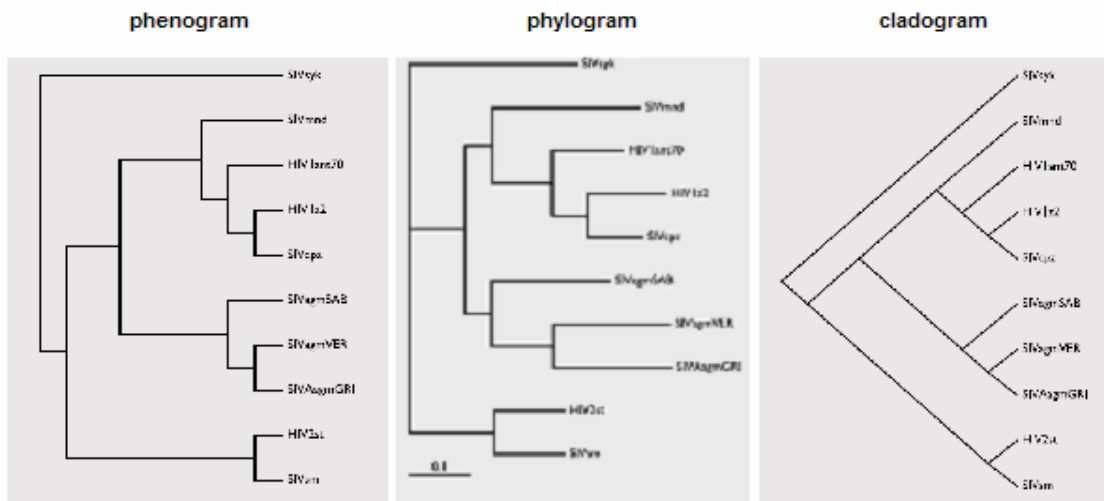
Figure 4: The three types of phylogenetic trees (Clewley 1998).

Finally, confidence measures play an important role in phylogenetics, especially when trees serve as the conceptual framework for the study of trait evolution (Alfaro et al, 2003). In order to be confident that a phylogenetic tree is accurate, the bootstrap method was introduced into the estimation of phylogenetic trees (Efron et al, 1996). The bootstrap is a computer-based technique for assessing the accuracy of almost any

statistical estimate (Efron et al, 1996). When applied to phylogenetic trees, the bootstrap assesses the "confidence" for each clade of an observed tree, based on the proportion of bootstrap trees showing the same clade (Efron et al, 1996). Also used to measure confidence, Bayesian analyses make inferences of phylogeny based upon the posterior probabilities of phylogenetic trees (Huelsenbeck and Ronquist, 2001). In the face of the widespread and growing use of Bayesian measure of confidence, its relationship with the bootstrap method is poorly understood, but both prove effective (Alfaro et al, 2003). It is important to make sure the data placed into the phylogenetic tree are accurate, therefore a measure of confidence is important in constructing phylogenetic trees.

Phylogeography

Phylogeography is a term first used in 1987 by John Avise (Avise, 2000). Phylogeography is the study of the forces regulating the geographical distribution of genealogical lineages, particularly those at the intraspecific level (Hare, 2001). It has been a course of study for a long time, but with the new advances in science and technology, such as the use of mtDNA, it has made it possible to trace the genealogical lineages across different geographical regions (Emerson and Hewitt, 2005). By sequencing a number of individuals across a species' range, it will help to make inferences about the geographic genetic diversity and evolutionary history of different populations and species (Knowles and Maddison, 2002).

Population genetics is the analysis of genetic relations between different populations over time. Evolution is the change in allele frequencies from generation to generation. The amount of genetic variation in populations plays a role in this process.

The forces that act on variation are genetic drift, mate selection, gene flow, and natural selection. Genetic drift is the tendency for the frequency of an allele to vary over time by random chance. Genetic drift decreases the genetic variation in a population. Mate choice can be either random or by selectionof particular traits. If the latter, the favored traits will increase in frequency while the unfavorable ones decrease. Gene flow is the movement of alleles from one population to the gene pool of another population. This increases the variability within a population and decreases the diversity between the two populations. Natural selection is a process in which traits that adds to an individual's fitness increase in frequency over time. With natural selection alone, the genetic variability will either remain the same or decrease.

Phylogeography is based largely on the coalescence theory. The coalescence theory is that any set of genetic sequences can be traced back to a single common ancestor, signifying a coalescent event (Rieseberg, 2000). The lines of ancestry from the genes or alleles can be used to form a gene genealogy by tracing back to the most recent common ancestor. MtDNA is most commonly used to track coalescence because it is easier to trace back, since it is passed on maternally and does not undergo recombination. It also factors in different evolutionary forces, like population expansion, population bottlenecks, genetic drift, mutation rates, natural selection and migration. These forces create patterns of coalescence times for genes (Beebee and Rowe, 2004).

The majority of phylogeographic research within animals is being performed using mitochondrial DNA (mtDNA) analysis. Also, plants are dominated by research with chlororplast DNA for similar reasons that mtDNA is used. MtDNA exhibits an increased amount of interspecific and intraspecific genetic variation relative to the

13

nuclear genome. This is because mtDNA has a higher mutation rate and unlike when a nuclear genome makes a mistake, mtDNA is unable to self correct its mistakes, since it does not have a DNA repair system. This makes it easier to trace back mtDNA back to an individual than nuclear DNA. MtDNA is passed on uniparently, from the mother, making it easier for researchers to trace back since it can be traced back exactly to one parent. Nuclear DNA is passed on from the father's and mother's sperm and egg cells. These cells undergo recombination during the formation of the gametes when crossing-over occurs in Prophase I. However, the Y-chromosome is being used for research, since it contains a section that does not undergo recombination. The effective population sizes of mtDNA and Y-chromosome are dependent on the number of males and females in the population respectively. Autosomal nuclear genes represent the entire population. Therefore, the effective population size will be smaller for both mtDNA and Y chromosomes. Due to having a smaller population, they are more susceptible to genetic drift. This makes the use of mtDNA useful, especially when these populations have no gene flow. By placing a strong effect on genetic drift, signs of divergent evolutionary pathways will be first recognized in mtDNA before appearing in nuclear DNA. (Halliburton, 2004). This research will either further collaborate or disprove the current inferences about the geographic distribution of genealogical lineages and could provide additional information on the evolutionary history of different species.

The dispersal of different haplotypes is strongly affected by physical barriers and demographic events. Physical barriers decrease the amount of gene flow into a population. This causes evolutionary factors, like genetic drift, natural selection, and mating selection, to act intensely on the population and increases the favored haplotypes,

possibly leading to allopatric speciation. Demographic events are incidents that lead to population changes. It is an increase or decrease on the total population. Some factors of this are birth rates, death rates, and sex ratio. Using the dispersal of different haplotypes, it will help researches make inferences on past events, such as population bottlenecks, expansions, and migration.

Tests of population structure and relative gene flow could improve with the use of maximum likelihood methods (Hare, 2001). The maximum likelihood method takes a model of sequence evolution and determines what the probability is that the observed sequence set would be produced (Beebee and Rowe, 2004). Models of sequence evolution are based on DNA sequencing and restriction mapping in which the phylogenetic relationships among their sequences can be inferred (Iizuka, 1989). Bayesian analyses determine the posterior probabilities, which are represented at the nodes of the tree. The values range from 0 to 1 and select a tree that is the best fit for the model under a specific model of nucleotide substitution. These methods can be used to estimate the divergence time and migration rate for two separate populations (Huelsenbeck et al., 2001). Both of these methods use the Markov Chain Monte Carlo algorithm and estimate the same factors under the same population model and assumptions. The two methods vary only in parameter proposal distribution and maximization of the likelihood function (Beerli, 2005).

Empirical Examples

Phylogeography is the study of the potential geographic explanations for the causes for differentiation among populations. The majority of these studies have used

mtDNA to show the variation among individuals. Buhay and Crandall (2005) performed a subterranean phylogeography study on freshwater crayfish in a cave assemblage. They looked at *Orconectes incomptus*, *Orconectes australis australis*, *Orconectes australis packardi,* and *Orconectes sheltie.* Using data from mtDNA 16s haplotypes, it was found that *O. a. packardi* is last common ancestor of the cave assemblage, and that the other species diverged from them. Because of a crayfish's stable underground environment and low dispersal rates, they have been able to persist and differentiate for millions of years. Their environment creates a geographic barrier that led to allopatric speciation. With neighboring range expansions and periods of isolation, it has led to increased variation within different species with the help of natural selection, genetic drift, and periods of gene flow and no gene flow. This is also seen with invasive species, where genetic diversity levels and population size builds up and recovers instead of leading to population bottlenecks (Tsutsui et al., 2000; Kolbe et al., 2004). An introduction of a large invasive population has a better possibility of establishing itself without undergoing a founder effect. Another study with troglobitic beetle species supported this. Sbordoni introduced 50 individuals into an isolated cave and after 30 years passed, the population had increased in size and genetic diversity (Sbordoni 1982). Finlay et al. (2005) found that *Cambarus tenebrosus,* a freshwater crayfish, with a stable habitat, for example subsurface, were found to have a large amount of genetic diversity. Because of their ability to live underground, it was a safe place for them during extreme habitat changes, like the glacial/interglacial periods. With the ability to survive on the surface and subsurface, an invasive species could be a potential danger to both surface and subsurface ecosystems (Finlay et al., 2006)

Levels of genetic variation can be used to distinguish one population from another. In a study about allozyme variation in *Procabarus clarkii*, genetic variation levels were far lower in populations from aquaculture ponds than those from wild populations (Busack, 1988). Those populations in the aquaculture pond were suggested to have high levels of inbreeding. Also, in another study, Fetzner found that the imported crayfish in the aquaculture ponds were having difficulty adjusting to the new habitat. This was creating a founder affect resulting in low genetic variance (Fetzner et al., 1997). Other factors contributing to low genetic variance and causes of population bottlenecks are habitat destruction, water pollution, competition with introduced species, low population numbers, and small size ranges (Taylor et al., 1996; Crandall, 1998; Horwitz 1990, 1994, 1995; Richardson et al., 1999).

METHODS

Crayfish Collection and Storage

     Crayfish were collected from three different sites. The three sites are located at Institute Pond, on the French River, and on the Blackstone River.
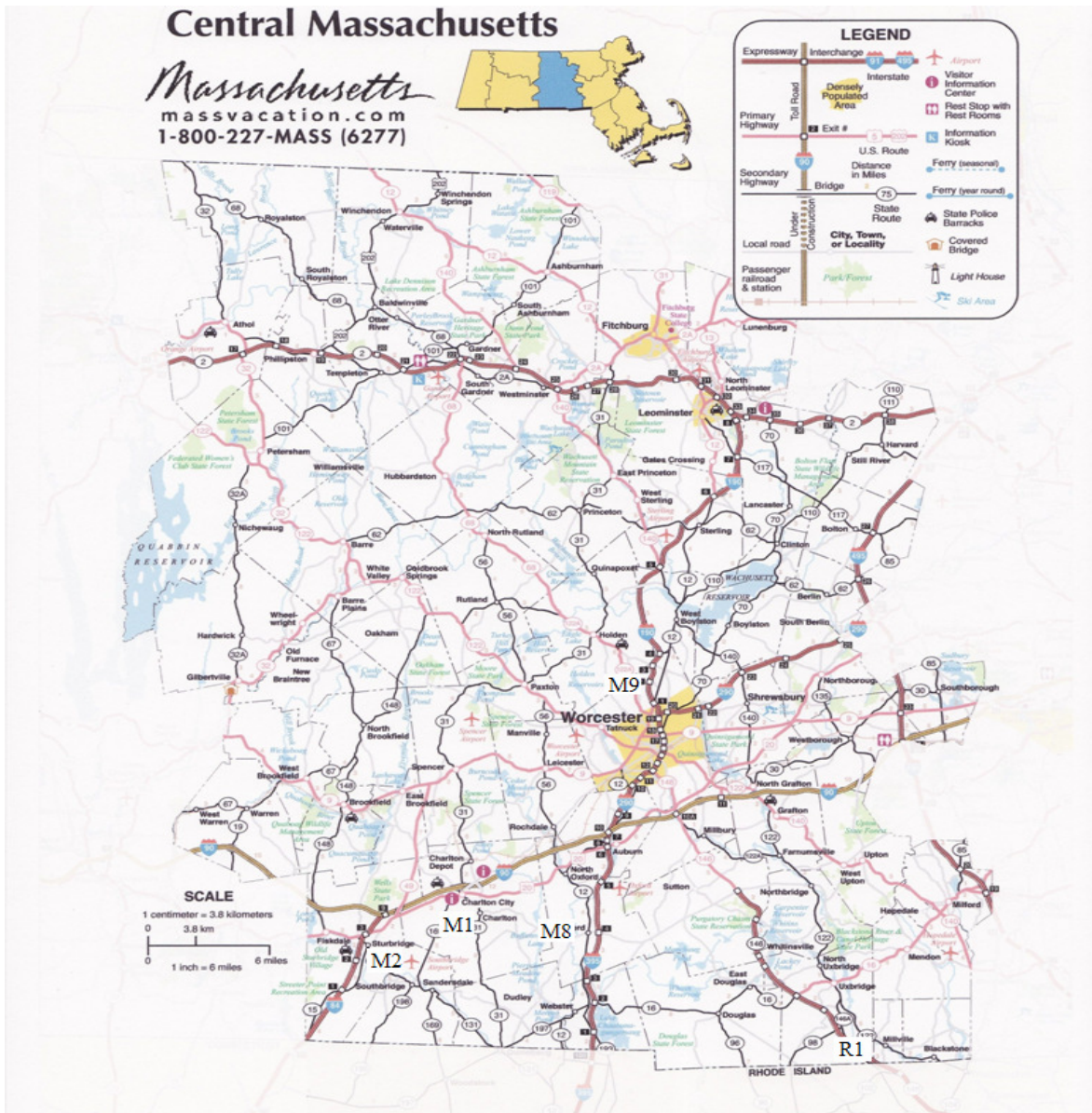

Figure 5: Map showing the placement the three sites, Institute Pond(M9), Hodges Village Dam(M8), Blackstone Gorge(R1), and the two sites from Gottardi and Anderson (2006); a manmade Charlton pond(M1) and Westville Pond(M2).

Institute Pond, also known as Salisbury Pond, is located next to Worcester Polytechnic Institute in Worcester, MA. The pond was created in 1834 when the Mill Brook was dammed. Water runs into the pond mainly from Mill Brook and Indian Lake but there is also storm water runoff (Lycott, 1989). The crayfish were obtained by trapping. Four traps were set ~10 feet from shore with either sardines or salmon in the trap as bait. Traps were located ~8 m from one another. The traps were cylinder shaped with one way openings on either side for the crayfish to enter.

Figure 6: The trap used for collection in Institute Pond

Figure 7: Institute Pond(M9)

Figure 8: Overview of Institute Pond

The traps were checked daily during the month of September, 2006. A total of 22 crayfish of species *O. virilis* were caught. Each day, the crayfish were collected into a 20 L bucket. They were then rinsed off with tap water and put into a 2 L bucket. The crayfish were then euthanized and preserved by being placed into a -20° C freezer.

The second site was from the French River at Hodges Village Dam in Oxford, MA. The French River starts near Leicester, MA and runs a total of 21 miles. In northeastern Connecticut it joins with the Quinebaug River (Department of the Interior, 2007; Bowles, 2007). The crayfish were collected by hand and net by turning over rocks or by use of a seine. A total of 30 *O. virilis* was collected. The crayfish were collected

into buckets with ~3 cm of water and were immediately returned to the lab after the collection where they were euthanized by freezing.

Figure 9: Hodges Village Dam: French River(M8)
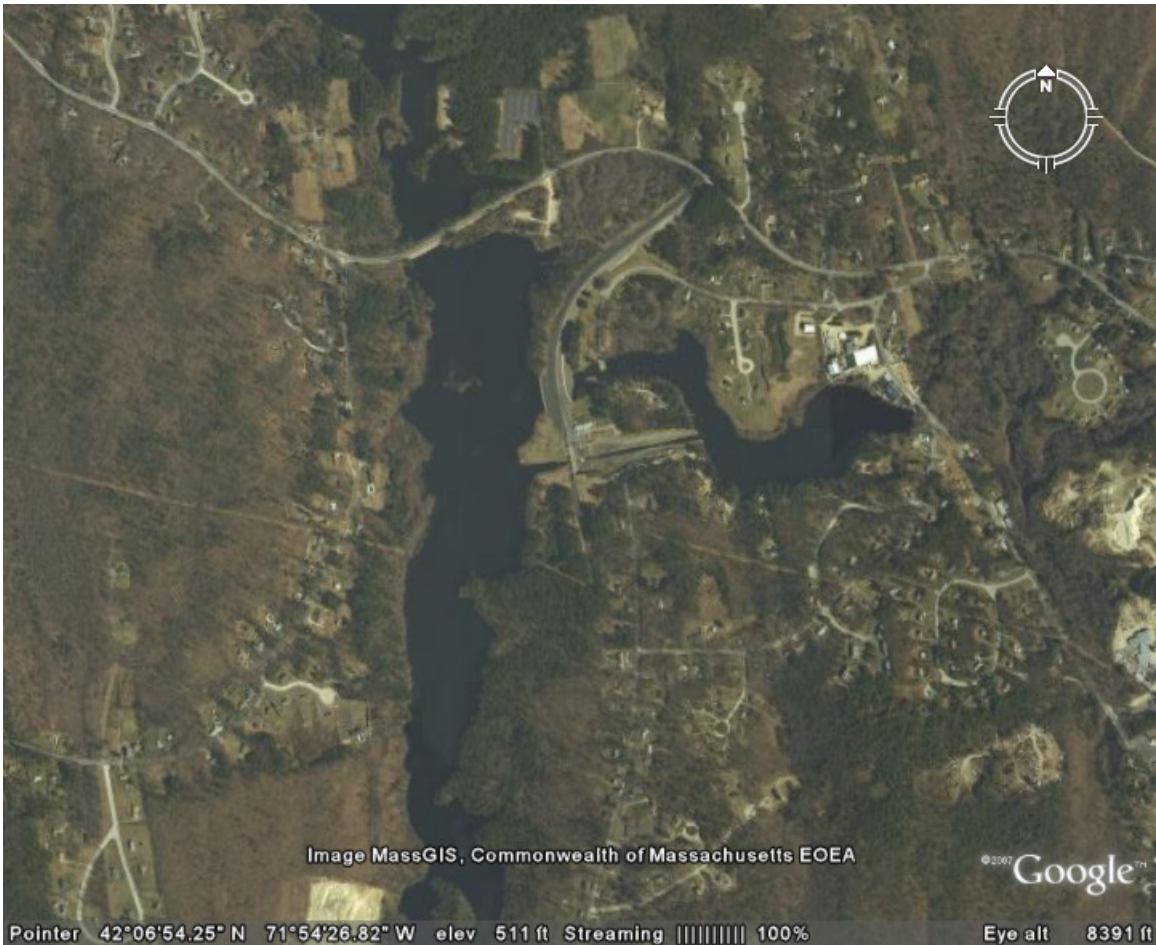
Figure 10: Hodges Village Dam, French River Overview

The final site was at the Blackstone Gorge in Rhode Island. The Blackstone River starts in Worcester with the flow of Mill Brook and Middle River (Kerr, 1990). It flows into Rhode Island and is a total of 43 river miles (Wandle and Phipps, 1984). A total of 30 *O. virilis* were caught. These crayfish were also caught by hand and net. They were then returned to the lab to be rinsed and euthanized.

Figure 11: Blackstone Gorge(R1)

Figure 12: Blackstone Gorge Overview

All the crayfish were labeled with a number-letter code depicting the site, species, and crayfish number (Table 1). Each crayfish was then examined, and the sex and carapace length were recorded. They were then placed back in the freezer until they were ready for DNA extraction.

Table 1: Site Information and Number of Samples

Site	Site Code	# of Samples	Longitude	Latitude
Institute Pond	M9	20	71°48'19.60"W	42°16'38.47"N
French River	M8	12	71°54'31.35"W	42° 6'54.62"N
Blackstone Gorge	R1	14	71°33'12.30"W	42° 0'53.25"N
Charlton	M1	19	71°59'16.17"W	42° 7'9.09"N
Sturbridge	M2	18	72° 3'36.47"W	42° 4'30.52"N

Additional data were included in this project from Gottardi and Anderson (2006). Their two sites were from a man-made pond in Charlton (M1) and from a site on the base of a dam at the East Brimfield Lake (M2). The East Brimfield Lake flows into the Quinebaug River, which the French River flows into at a site in Thompson, CT, but its conjunction is too far downstream from the French River site and Sturbridge site to have any correlation. Both the Blackstone and Institute Pond are part of the same water shed with the Blackstone site being approximately 25 miles downstream of the Institute Pond site.

DNA Extraction

DNA was extracted from each of the labeled crayfish from the three sites using the Solid Tissue Protocol in Gentra System's Puregene: Genomic DNA Purification Kit. Using this kit allowed a fast, easy, and inexpensive way to extract and purify the crayfish DNA. Muscle tissue was extracted from either a cheliped or from the abdomen depending on the size of the crayfish. The amount of tissue extracted was approximately 5-10 mg, measured qualitatively because the kit was insensitive to small differences in starting

material (Mathews, personal communication). The remainder of the crayfish was refrozen in order to preserve the DNA in case this step had to be redone. The tissue was placed in a labeled microfuge tube with cell lysis solution (300μL). The cell lysis solution caused the muscle cells to swell and then burst, allowing the cellular components to mix into the solution. Then 1.5 μL Proteinase K solution was added to each sample. Proteinase K broke down the peptide bonds linking amino acids from each protein through a method called hydrolysis. Each sample was inverted 25 times to mix the solution and was incubated at 55 ° C overnight in order to dissolve the tissue. A volume of 1.5μL of RNase A solution was then added to the cell lysate. The RNase A destroyed the RNA found in the cell lysate by cleaving the 3' end of the C and U residues and did not affect the DNA. Then 100 μL of Protein Precipitation Solution was added to each tube, and the tubes were vortexed, then centrifuged at 13,000-16,000 g for 3-5 minutes. This caused the protein to precipitate and produce a protein pellet at the bottom of the tube. The supernatant was removed to another microfuge tube along with 300 μL of 100% isopropanol. This causes the DNA to precipitate out of solution. 0.5μL of 20mg/mL glycogen was added to the mix to increase the amount of DNA pellet since glycogen is a DNA carrier. This mixture was centrifuged at 13,000-16,000 g for 1-5 minutes, causing the DNA to form a pellet. The DNA was then separated from the supernatant and was rinsed with 70% ethanol to wash the pellet. It was centrifuged again under the previous conditions for 1 minute and supernatant was removed to isolate the DNA. The DNA pellet was placed in 50 μL of DNA Hydration Solution. It was left at room temperature overnight with periodical tapping and then placed in the freezer for storage and further tests.

Gel Electrophoresis

In order to actually determine that the product from the DNA extraction and PCR experiments contained DNA, each sample was run through a 1% agarose gel. The 1% gel was prepared by adding 1g of agarose into 100 mL of 1X TAE buffer. It was heated and 2 µL of 10 mg/ml of ethidium bromide was added to the mix. The ethidium bromide intercalates with the DNA and will fluoresce a orange-red when it is exposed to UV light. The gel was poured and left to cool to room temperature. The gel then cooled and the agarose formed pores proportional in size to the amount of agarose in the solution. The wells of the DNA are towards the negative end. This is because DNA has an overall negative charge and is attracted through the pores by the positive end of the gel. The gel was run with 50 mg of lambda DNA standard to estimate the concentration and size of the DNA. They were loaded with a blue dye to track its progress across the gel. The gel was run with an output level of about 100 volts until the tracking dye was approximately ¾ the way though the gel. It was subsequently viewed under UV light.
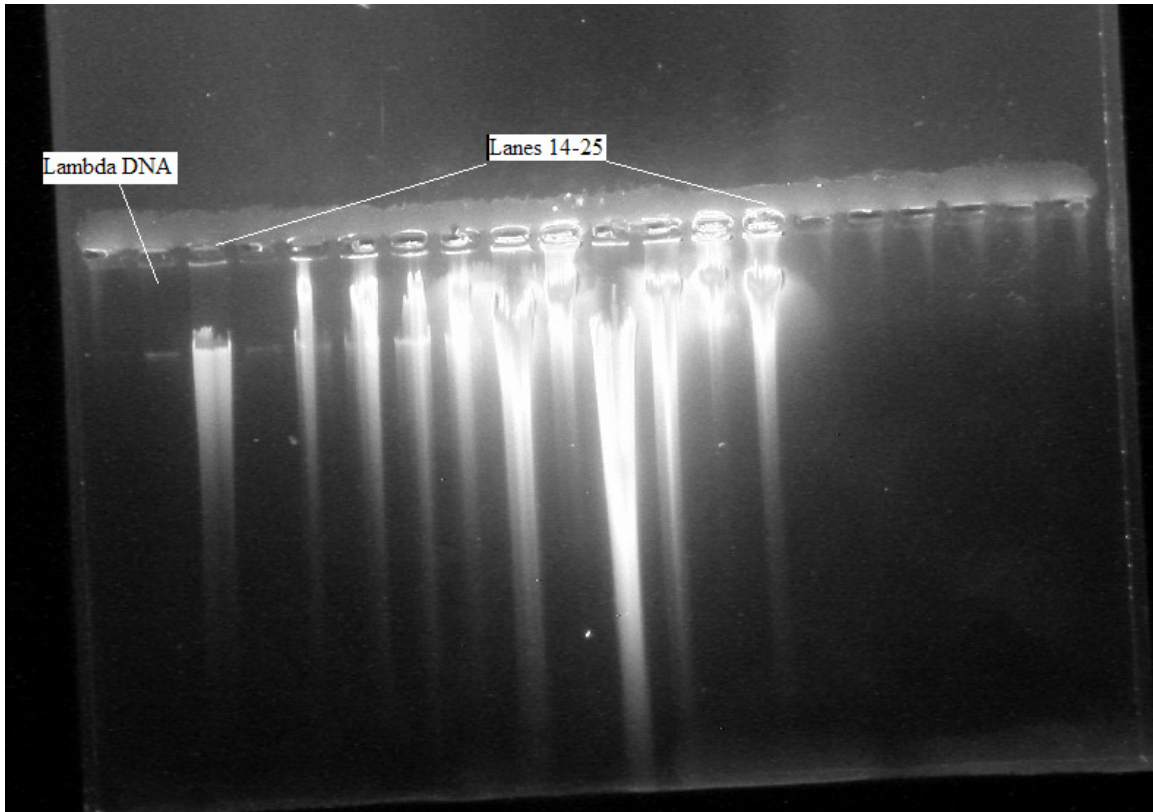
Figure 13: Picture of the gel for M9 samples 11-22 following the DNA extractions. Was run with Lambda DNA to qualitatively determine the concentration based on brightness.

Polymerase Chain Reaction

Polymerase chain reaction, known as PCR, is a procedure that allows one to amplify a gene or locus by making millions of copies of the desired DNA sequence it lies in. It was developed by Kari Mullis in 1983 and has since offered a fast and easy way to amplify DNA (Saiki et al., 1988). The basic components of PCR are a template of the DNA of interest, primers, Taq polymerase, deoxynucleotide triphosphates, and a buffer solution. The primers must be 5' to 3' and be complimentary to the DNA of interest, flanking the gene of interest. Taq polymerase is a polymerase from the bacterium Thermus aquaticus, used because of its ability to withstand the temperature conditions of

the reaction without being denatured (Saiki et al., 1988). The dideoxynucleotide triphosphates are free flowing bases that will be used to make up the product of PCR.

PCR takes place in cycles. Each cycle is comprised of three main steps. The first step is the denaturing step, which separates the double stranded DNA into single stranded DNA. The temperature is raised to 94-95° C, which causes the hydrogen bonds between the strands to be interrupted. This step is usually 20-30 seconds, but for the initial step it is performed for 1-9 minutes. Following this step is the annealing step, where the temperature is lowered to 50-64° C, depending on the primer. During this step the primer anneals to its complimentary section of the DNA, and the complementary DNA cannot rehybridize, because of the large concentration of the primer. This step is usually 20-40 seconds. The last step of the cycle is the extension step. It is performed typically at 72° C because Taq polymerase is active between 70-74° C. The time of this step is based upon the size of the segment of DNA being copied, but typically 1000 base pairs can be polymerized in a minute. This cycle is repeated until the amount of DNA amplified is sufficient (Voet et al, 2006).

For our reactions, the DNA being amplified was the mitochondrial 16S ribosomal RNA gene. The two primers that were used were 16S-1472 (5'-AGATAGAAACCAACCTGG-3') and 16S-L2 (5'-TGCCTGTTTATCAAAAACAT-3'). The starting volume of the PCR mixture was 20 ųL each. The components were 20 ng of the genomic DNA containing the16S gene, 1.6μL of 10μM of each primer, 1.6μL of 2mM of each dNTP, a buffer concentration of 1X Pico Maxx Reaction Buffer, 11.4 μL of autoclaved deionized water, and 0.2 units of Taq polymerase. The 16S gene was roughly 20 ng, because the concentration of the DNA was determined qualitatively based on its

comparison to the standard. A fraction of the DNA extraction product was diluted initially to 10 ng/μL, the rest was put away for storage in case of any error or need of repetition. The initial denaturing step was at 95° C for 2 minutes, but every other following cycle ran for 30 seconds. The annealing cycle was at 48° C for 30 seconds. The extension step ran for 1 minutes at 72° C and the steps were repeated for a total of 40 cycles. During the last extension cycle, it was run for 10 minutes. This was to make sure that all the single stranded DNA had polymerized. After all was done, each sample underwent gel electrophoresis to seek a positive result from PCR. After PCR, gel electrophoresis was performed on all the samples. All the steps were the same except a higher concentration of agarose was used after PCR because the size of the DNA had been reduced after PCR to a length of approximately 550 base pairs and the pore size needed to be smaller.
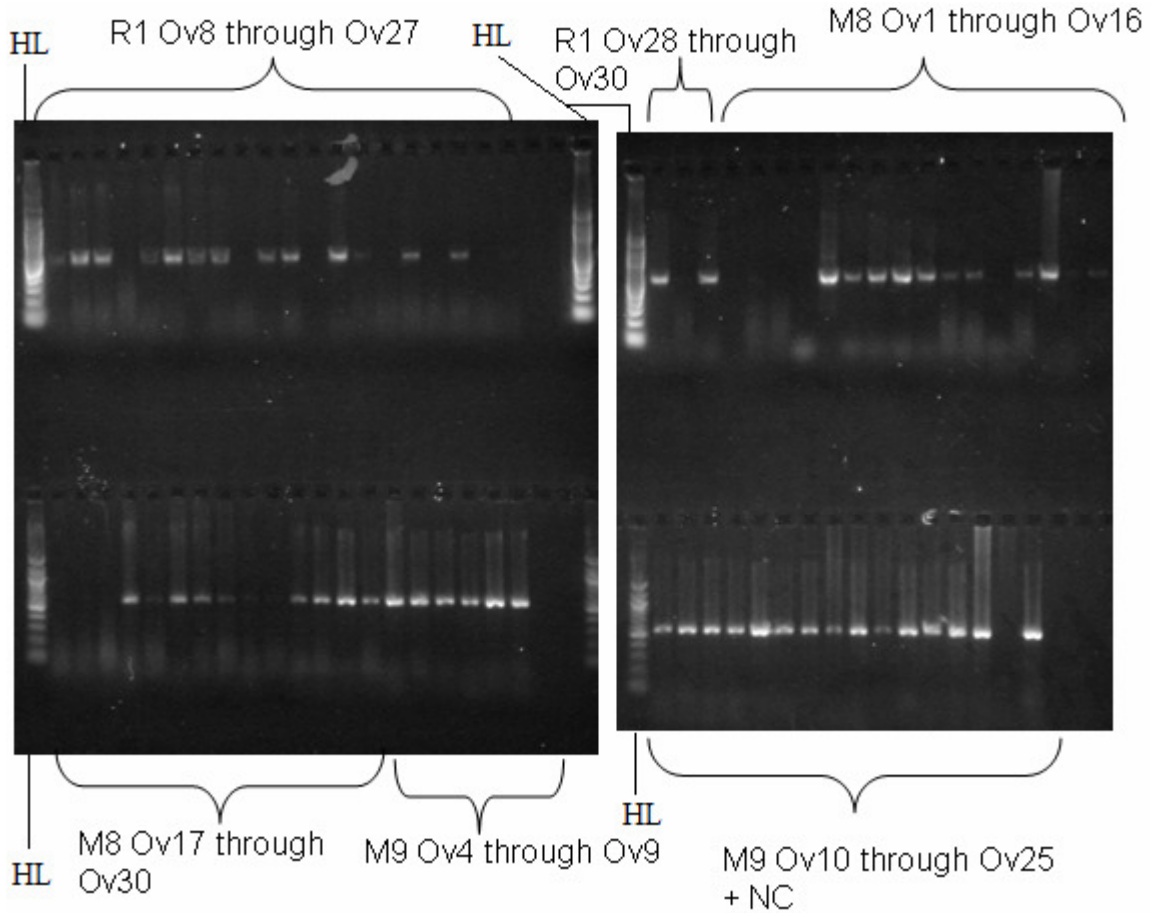
Figure 14: Picture of Gel Post PCR: Lanes represented by HL are the ladders.on the very left of both photos and on the very right of the left photo ladders. Bands shown with the sample lanes positive for PCR product show a ~550 base pair band. The lane with NC is a negative control.

DNA Sequencing

The idea of DNA sequencing is to determine the exact positions of the nucleotide bases in a strand of target DNA. The world of DNA sequencing was changed by the onset of Sanger sequencing which utilized chain termination techniques and gel-electrophoresis to shorten the time it takes to sequence DNA and to also sequence more base pairs than allowed by previous methods. The Sanger sequencing method uses the

properties of the 2', 3'-dideoxy and arabinonucleoside analogues of the normal deoxynucleoside triphosphates, which cause chain termination of DNA polymerase (Sanger et al, 1977). In original Sanger sequencing, one base, thymidylic acid, or thyamine (T), was targeted. When thymidylic acid (dT), was replaced by 2',3'-dideoxythymidine triphosphate (ddT) the DNA polymerase, due to the fact that ddT is unable to bond to more nucleotides, was unable to continue the polymerization of the new DNA strand, and the chain was terminated (Sanger et al, 1977). In a mixture with a high percentage of dTs compared to ddTs the polymerase creates DNA strands of varying lengths (Sanger et al, 1977). When a primer and template are incubated with DNA polymerase in the presence of the mixture stated above and with the other bases, A, C, and G, the result is a mixture of DNA strands with the same 5' end and with ddT residues at the 3' ends (Sanger et al, 1977). Due to the sheer amount of ddTs and dTs present every dT position has been substituted by a ddT, therefore marking each position in which a dT occurs. In order to determine the position of the ddTs, the resulting mixture is run through a gel electrophoresis. This process separates the DNA strands by size, while drawing the negative charge of the DNA towards the positive node. The gel, treated with ethidium bromide, which binds the DNA and fluoresces under UV light, then displays dark bands at the position of each ddT. When the same template is run with dd derivatives of A, C, and G, the entire sequence can be obtained (Sanger et al, 1977).

Based upon the idea of Sanger sequencing which rely on the principles of chain termination, automated sequencers with laser detection systems have been built to increase the speed of DNA sequencing based on the same These methods are dye-terminating or dye-primer reactions. The dye terminating reactions occur in the same

tube, where a DNA polymerase draws from a mixture of dideoxynucleoside triphosphates (ddNTPs) which stop polymerization, and deoxynucleoside triphosphates (dNTPs) which are the standard bases used in DNA polymerization (Lee et al, 1992). Once the polymerization has completed, each of the ddNTPs has been bonded by a different fluorescent dye. The mixture is run through gel electrophoresis and when the gel is put into UV light, the dyes fluoresce in different colors and the position of each nucleotide can be distinguished (Lee et al, 1992). Another method is dye-primer sequencing, which relies on the principle of binding a primer with a dye, and running one primer at a time for each ddNTP. The total of all four ddNTP primer reactions are pooled and separated by gel electrophoresis and the sequence of the DNA can be determined (Lee et al, 1992). With the help of computers, these reactions are able to sequence DNA quickly.

The target DNA for this project was the 16S gene in the mitochondria of each individual crayfish collected. During PCR primers for both the forward and reverse strands were bound and polymerized. Both forward and reverse strands were prepared to be sequenced. The following prepatory sequencing reactions were performed on the PCR product. The reactions were done in 10 µl volumes which consisted of 2 µl PCR product, 1 µl of [3.3µM] each ddNTP primer, 1.5 µl 5X buffer and 0.5 µl of BigDye 3.1 terminator mix (ABI) and the remaining volume was filled with water. This mixture was placed into the thermal cycler and run using the specified program. The program for sequencing is 40 cycles at 96° C for 10 seconds, 50° C for 5 seconds, and 60° C for 4 minutes. This reached a final temperature of 15° C for 20 minutes after the 40 cycles finished, and then the tubes were kept at 4° C indefinitely.

The PCR product then needed further preparation for DNA sequencing. Volumes of 25 µl of 90% ethanol and 1 µl of [125mM] sodium acetate were added to the remaining product. The tubes were inverted to mix and were placed in a -20° C freezer for 20 minutes. The tubes were then placed into a centrifuge and spun at 3000Xg for 15 minutes. The supernatant was removed and the tubes were placed uncapped back into the centrifuge upside-down for approximately 5 seconds two times to dry the product. The remaining pellets were washed with 40 µl of 70% ethanol and then centrifuged at 3000Xg for another five minutes. The product was then dried in the same fashion as stated above. After being dried the pellets were resuspended in 10 µl of deionized water. The preparations for the automated sequencing were complete, and the products were sent to the Biotechnology Resource Center at Cornell University, in Ithaca, New York, for automated sequencing on an ABI 3730xl capillary DNA sequencer.

Data Analysis

When the results of the sequencing reactions came back, the files were loaded into the LaserGene (DNASTAR) and other analytical computer programs. The first program used was FinchTV (Geospiza). This program was used to see the DNA sequencing results and determine which sequences returned useable data for analyzing. Once the sequences were narrowed down, the next program used was SeqMan, which allowed us to determined consensus information by visual inspection of the forward and reverse fragments from each crayfish. The sequences were then all reinspected in FinchTV to determine the exact nucleotides in certain positions where ambiguity still remained. After closer examination with the FinchTV program, BioEdit v. 7.0.4.1 (Hall 1999) was

used to change the ambiguous nucleotides to the correct ones. In addition to the sequences obtained from the three sites of this project, sequences from Gottardi and Anderson (2006) were added to allow for a better analysis of all the data. Once all the bases positions had been confirmed, the sequences were all aligned by the ClustalW method in Megalign (DNASTAR). The outgroup sequence that the data was run against was that of *Orconectes limosus*.

To prepare the data to be run in Mr.Bayes, MrModeltest (Nylander, 2004) was used to determine an evolutionary model that best fit the data. This was done because all phylogenetic methods make implicit or explicit assumptions about the process of DNA substitution and Huelsenbeck and Rannala (1997) stated that maximum likelihood (ML) provides a rational method for choosing substitution models for phylogenetic analysis through the use of likelihood ratio tests (Largiader et al., 2000). MrModeltest executes likelihood ratio tests and outputs a DNA substitution, evolutionary model that best fits the input data (Castro and Dowton, 2007). MrModeltest (Nylander, 2004) showed that the General Time Reversible model was best for use in MrBayes (Huelsenbeck and Ronquist, 2001).   The data was ready to be analyzed in Mr.Bayes.

Only unique haplotypes were included in the Bayesian analysis.   These haplotypes were assigned by creating a pairwise identity matrix in BioEdit, and importing that information into a Microsoft Access Database. The new data set of 10 unique haplotypes and 1 outgroup was then run through the Mr.Bayes (Huelsenbeck and Ronquist, 2001) program.

Once the data was input into the Mr.Bayes program, it was run against a General Time Reversible model and gamma distributed rate variation. Markov Chain Monte Carlo

(MCMC) analysis was run for 1,000,000 generations, and 15,000 generations were discarded as burn-in.

These MCMC results were saved as a ".con" file, which was opened in TreeView. The TreeView program displayed the results of the MrBayes program in the form of a phylogenetic tree, which has been detailed and displayed in the results section.

After the final alignment was obtained, the sequence information was input into Arlequin v. 3.11 (Excoffier et al., 2005), to investigate molecular variability in the data set. This program was also used to calculate the nucleotide and haplotype diversities. Arlequin was used to determine nucleotide diversity, which quantifies the polymorphisms between individuals within a population, the 5 locations the crayfish were collected from in this case, as developed by Tajima (1983). Nucleotide diversity is defined as the probability that two randomly chosen nucleotides are different (Halliburton, 2004). Haplotypes are defined as a unique sequence of linked genetic markers, which are composed of nucleotides (Halliburton, 2004). Mitochondrial DNA is passed on uniparentally from the mother, this ensures that each of the sequences, which are unaltered by genetic recombination, are the same for each offspring except when a mutation occurs. In addition to that, the mtDNA does not have a repair mechanism, so when mutations occur during DNA replication new haplotypes can arise. Haplotype diversity compares each set of nucleotides as a whole, whereas the nucleotide diversity compares random, single nucleotides from the mtDNA sequences of the crayfish. The test of nucleotide diversity may show insertion/deletion mutations that do not lead to new haplotypes, whereas the haplotype diversity is a measure of how diverse a population is by looking at the dispersal of the haplotypes within the population and between other

populations.   Haplotype diversity was also calculated and used to display the polymorphisms that arose within these 5 populations according to procedures described by Nei (1997).   An analysis of the molecular variance (AMOVA), as described by Excoffier et al. (1992), was run in Arlequin.   The AMOVA is based of the principles of the analysis of variance (ANOVA) (Halliburton, 2004).   The ANOVA partitions the variance of allele frequencies in the whole population into components representing variance among individuals within subpopulations and variance among subpopulations (Halliburton, 2004).   The AMOVA simply allows the techniques of the ANOVA to be applied to DNA sequences instead of just on allele frequencies and heterozygosities (Halliburton, 2004).   The AMOVA sets up a distance matrix of haplotypes and this data is run through ANOVA, which provides information of variance within population and among populations, in addition it also computes F_{ST} values, which represent heterozygosity deviations of the individuals in a population from the group of all the individuals of all the populations (Halliburton, 2004).   Along with output F_{ST} values, P-values are output, which evaluate the significance of the F_{ST} values.   The P-values range from 0-1, a P-value approaching 1 represents insignificance in the output F_{ST} value, whereas a P-value approaching 0 represents large significance of the F_{ST} values output. For this experiment, $\alpha=0.05$ such that any F_{ST} with a P value under 0.05 could be considered to be statistically significant.

RESULTS

The results of DNA sequencing gave us 46 usable sequences from the novel collection sites M8, M9, and R1. In combination with the data from Gottardi & Anderson (2006), we had a total of 82 sequences. The Institute pond site gave the highest success rate for usable sequences, with a total of 20 sequences from 22 attempted. The French River produced a total of 12 usable sequences from 30 attempted. The Blackstone Gorge produced 14 usable sequences from 30 attempted. The unsuccessful sequences were either due to an unsuccessful DNA extraction or PCR amplification, or resulted in an unreadable or incomplete sequence. The following table shows the different haplotypes found, the size of each haplotype, and the source populations for the haplotypes.

Table 2: Haplotype Data: Number of individuals with each haplotype and source locations

Haplotype	Number of Individuals	Location
	1	Outgroup
O. virilis haplotype 1	41	M8 (10), M9 (20), R1 (11)
O. virilis haplotype 2	1	R1
O. virilis haplotype 3	1	R1
O. virilis haplotype 4	1	R1
O. virilis haplotype 5	1	M8
O. virilis haplotype 6	22	M1 (6), M2 (15), M8 (1)
O. virilis haplotype 7	12	M1 (12)
O. virilis haplotype 8	1	M2
O. virilis haplotype 9	1	M2
O. virilus haplotype 10	1	M2

The three most frequent haplotypes found were haplotypes 1, 6, and 7. Haplotypes 2, 3, 4, 5, 8, 9, and 10 were found in only a single individual.

Figure 15: Bayesian phylogenetic tree displaying the relationships among the 10 different unique haplotypes of the *O. virilis* from the 5 sites. *O. limosus* was used as an outgroup. The number in parentheses represents the sample size (number of individuals) of each haplotype. To the right of each haplotype is a pie chart showing the relative representation of various collection sites for each haplotype. Numbers next to nodes refer to the posterior probabilities.
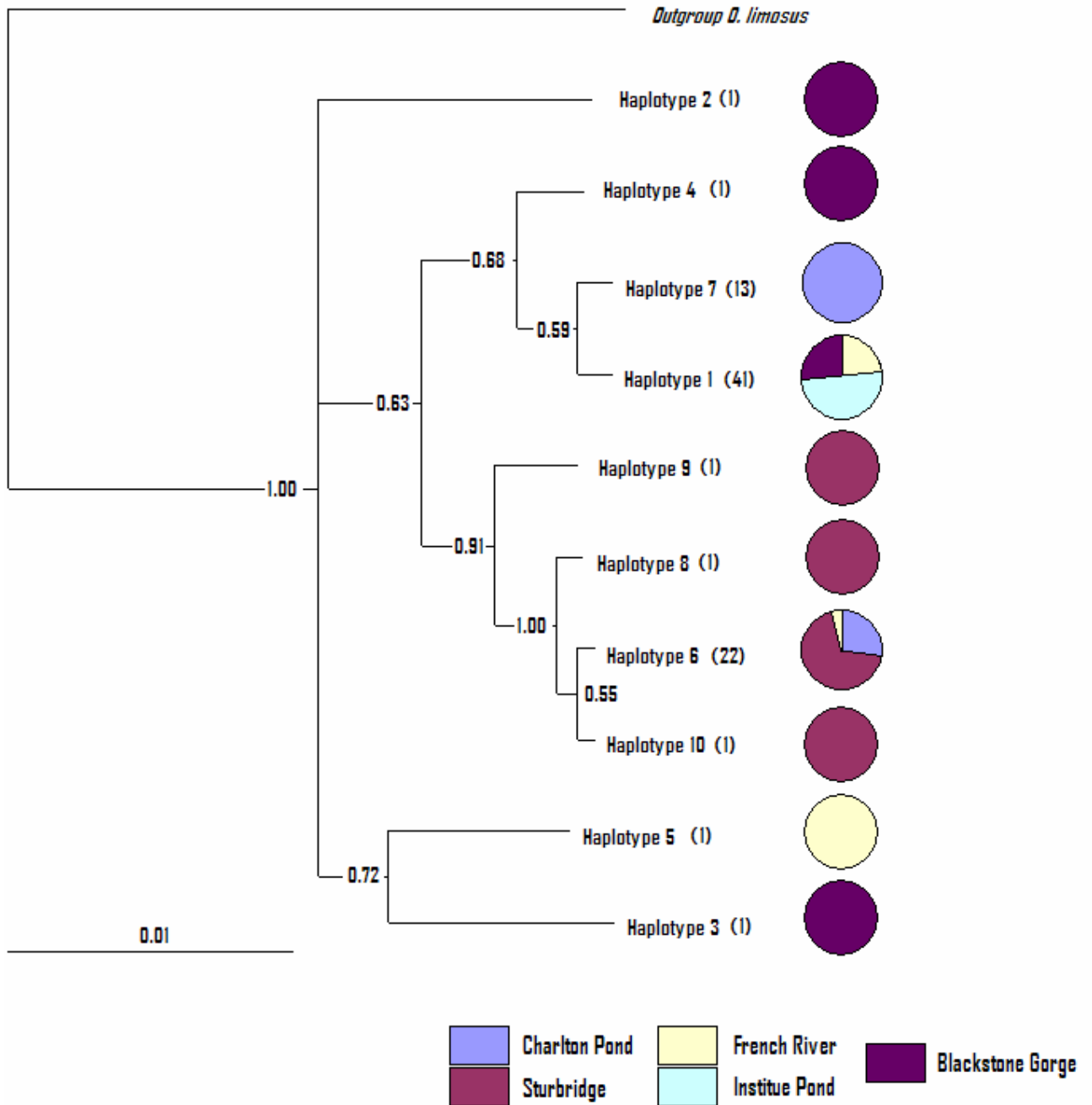
Figure 15 shows the phylogenetic tree of the 10 unique haplotypes. Using the posterior probabilities, the tree displays 3 different clades among the 5 populations. The clade containing haplotypes 1, 4, and 7 and the clade containing haplotypes 6, 8, and 10 represent the majority of the data. The location of the samples within the clades shows evidence that some of the populations may be related due to sharing the same water shed.

Table 3: Intra-population analysis of *O. virilis* populations

Site Code	Mean number of pair wise differences	Nucleotide Diversity	Haplotype Diversity
R1	2.938+/-1.637	0.006+/-0.004	0.396 +/- 0.159
M8	1.012+/-0.729	0.002+/-0.002	0.318 +/- 0.164
M9	0.000+/-0.000	0.000+/-0.000	0.000 +/-0.000
M1	0.335+/-0.350	0.001 +/- 0.001	0.471 +/- 0.082
M2	1.425+/-0.912	0.003+/-0.002	0.314 +/- 0.138

Table 3 above shows the intra-population analyses for the 5 sites. Four of the sites show little nucleotide diversity while the Institute Pond site shows none. The mean number of pairwise differences for the five sites ranged from 0.335 to 2.938. The haplotype diversity for the five sites showed little diversity, ranging from 0 to 0.471.

Table 4: Results of analysis of molecular variance on the 16S data; df= degrees of freedom; SS= sum of squares.

Source of variation	df	SS	Variance components	Percentage of variation
Among populations	4	62.31	0.91	56.24
Within populations	77	54.74	0.71	43.76
Total	81	117.05	1.62	
Fixation Index	F_{ST}: 0.562, P<0.00001			

Table 4 shows the results of the AMOVA for the five populations. The F_{ST} value equals 0.562 (P<0.0001). The P value signifies that the data is realistic. The percentage of variation among the populations(56.24) was found to be more than the percentage of variation within populations(43.76).

Table 5: Fst and P-value Pairwise Matrix: F_{ST} **values are below the diagonal and P-values are above the diagonal.**

	R1	M8	M9	M1	M2
R1		0.711	0.072	<0.0001	<0.0001
M8	-0.010		0.171	<0.0001	<0.0001
M9	0.036	0.045		<0.0001	<0.0001
M1	0.382	0.445	0.66		<0.0001
M2	0.653	0.79	0.95	0.564	

Table 5 shows the F_{ST} between the different populations. The data shows high F_{ST} values between all the sites that have a significant P-value. The different values of the pairwise F_{ST} appear to have a geographic relationship.

DISCUSSION

The overall F_{ST} value is ~ 0.5. This suggests that the crayfish from each location are highly differentiated from each other. An F_{ST} value of ~0.5 is fairly high. An F_{ST} value of 0 shows that the genetic diversity in each sub population is very similar to the genetic diversity of that of the combined sub populations, whereas an F_{ST} value as it approaches 1 shows high genetic differentiation within each sub populations from the overall population. The fact that these populations have an F_{ST} value of ~0.5 shows that the group of sub populations is highly differentiated from each other as a whole.

The evidence of founder effect can also be seen in the data. The Institute Pond Site shows all crayfish having the same identical haplotype. This suggests that each crayfish has a very recent last common ancestor due to the fact that each of their mtDNA sequences are exactly alike. This evidence shows that one female has started this entire population, which is very much that of a founder effect.

As stated above, the Institute Pond site has zero genetic diversity in both haplotype and nucleotide diversities, however, the Blackstone Gorge and French River have nucleotide diversities of 0.006008 and 0.002066, and haplotype diversities of 0.3956 and 0.3182 respectively. The haplotype and nucleotide diversities for the Charlton Pond and Sturbridge sites are 0.471 and 0.314, and 0.001 and 0.003 respectively. The nucleotide and haplotype diversities that resulted demonstrate average diversity when compared to data obtained from crayfish in France by Grandjean and Souty-Grosset (2000). The mean nucleotide diversity was 0.0035 and the haplotype diversity ranged from 0.05 to 0.75 (Grandjean and Souty-Grosset, 2000). This correlates well with our data, and shows that the crayfish are of average nucleotide and haplotype diversity.

The data also shows 7 singleton haplotypes, 3 in R1, 1 in M8, and 3 in M2. Singleton haplotypes in this discussion mean a crayfish in a population, or a site, that have a haplotype unidentical to any other crayfish of their population. Of the 82 individual crayfish, only 7 showed singleton, unique haplotypes across 3 out of 5 sites. Three haplotypes contain the majority of the individuals, while one haplotype, haplotype 1, contains nearly half (41 individuals). All 20 individuals from the Institute Pond site were the same haplotype, haplotype 1. The haplotype diversities of these crayfish from each site, besides the genetically non-diverse Institute Pond, are average in comparison to other crayfish studies. This suggests that the haplotype occurrences are average which is peculiar and not completely supportive of our hypotheses.

From the data, it appears that the three sites collected from this year have very little interaction with the two sites added from last year. Only haplotype 1 was observed from Institute Pond. The majority of the crayfish from Blackstone Gorge and the French River are also haplotype 1. However, these sites had some additional singleton haplotypes. A peculiar observation from the French River was also made. It was composed of 12 individuals, 10 haplotype 1, 1 haplotype 5, and 1 haplotype 6. The individual in haplotype 6 was accompanied by 21 other crayfish, all from either Charlton Pond or the Sturbridge site.

The phylogenetic tree constructed from our data also presents an interesting situation. Most of the crayfish are segmented into two major clades. The two minor clades are most likely a part of one of the two clades, but sample size was small, and the posterior probabilities are also not overly significant, therefore they appear as smaller clades and are not incorporated as we believe they should be. The two major clades are

45

interesting because of how the crayfish are divided into them. The second major clade is comprised of nearly all Sturbridge site crayfish with some Charlton Pond crayfish and one French River crayfish. The first major clade includes all 20 Institute Pond crayfish, 13 Charlton Pond crayfish, and some French River crayfish. This is peculiar because as opposed to Institute Pond being all one haplotype, this shows that the crayfish in the Sturbridge site have all genetically diverged from the rest of the crayfish. The crayfish from the French River that is included into this group is even more peculiar, however the Charlton Pond is in close geographic proximity to the Sturbridge site, therefore being more understandable as to how the same haplotypes, though diverged appear in both sites. In addition, the pairwise F_{ST} and p-values table demonstrates an F_{ST} value of ~0.5 between the Charlton site and the Sturbridge site. This shows high differentiation between the two and a low average dispersal of diversity throughout. This suggests little interaction between the two sites.

Of the 3 sites collected this year, the pairwise F_{ST} and p-values table, a negative F_{ST} is actually seen between the Blackstone Gorge and French River sites. This suggests that the heterozygosity for each subpopulation is greater than the heterozygosity expected for the total of all the populations. This does not correlate with any of the other data, and is also refuted by a very high P value of > 0.7. Other interesting results from the pair wise F_{ST} and P-value table is a relatively low F_{ST} value between the Institute Pond sites and the French River and Blackstone Gorge sites. These F_{ST} values are also supported with relatively low P-values.

The Institute Pond site and the Blackstone Gorge are a part of the same water way about 25 miles apart, the Institute Pond site being upstream from Blackstone Gorge.

46

There are no dams between the two sites (Kerr, 1990). Due to the geographic and geologic relationships between the two sites, it is not illogical to assume that some crayfish from the Institute Pond site would be able to travel downstream to the Blackstone Gorge site. This is shown in the pairwise F_{ST} and P values table, which shows and F_{ST} value of about 0.03 between the Blackstone Gorge and Institute Pond with a relatively low P-value of ~0.07.

The percent of variation data is also very alarming. The percent of variation within human populations found by Cavalli-Sforza (2005) was between 93-95%, whereas the variation among the populations was between 5-7%. Our data shows a variance of 54% among populations, and 46% within populations. Other AMOVA results from marine amphipods, *Gammarus locusta*, showed variation among regions to be 1.3%, among populations within regions to be 3.8%, and variation within populations to be 94.8% (Costa et al., 2004). Further evidence from Mathews (2006) shows variance analysis of 4 clades of snapping shrimp. The variation within populations for each clade was 99.35%, 76.39%, 81.42% and 23.63% (Mathews, 2006). The clade with 23.63% was found to be divergent hence lower variation within the population than among populations (Mathews, 2006).

Our hypothesis of a founder effect was found, however the average haplotide diversities and nucleotide diversities can not support the fact that more than Institute Pond has undergone a founder effect. Also, our hypothesis of low genetic diversity can not entirely be supported either due to the average levels of haplotype and nucleotide diversities. This presents an interesting possibility. The AMOVA analysis correlates with our phylogenetic tree in the fact that two major clades were produced and the

variation of our crayfish within populations was only 46%. When compared to several other species, this is very low. However, a low percent of variation within population could suggest divergence. This gives rise to a new hypothesis that maybe the crayfish are two different species, closely related, that are phenotypically indistinguishable when collecting them. These two crayfish may or may not have begun hybridizing, but the mtDNA phylogenetic tree shows clear divergence of two clades. The F_{ST} values and pair wise P-values also show that the Sturbridge site is genetically varied from the all the other sites, with the lowest F_{ST} value being ~0.5. This suggests that the genetic diversity in the Sturbridge site may result from the presence of two different species which have diverged, but possibly did not reproductively isolate or are hybridizing.

Conclusions and Future Directions

The data shows that the species *Orconectes virilis* could be invasive to the water systems of Massachusetts, and evidence of a founder effect has been observed. However, average levels of nucleotide and haplotype diversity suggest a perplexing new hypothesis. In order to further investigate the situation of *Orconectes virilis* in Massachusetts, more samples would need to be taken, and further research done. This research would need to include a broader range and more sites geographically proximal to the Sturbridge site to investigate the inhabitance of the species. With all this data, a determination on the state of *Orconectes virilis*, and/or another species if found can be more clearly made.

If the species is indeed invasive, which is the current assumption, the cause of invasion should also be investigated. At least three different mechanisms may play a role in the displacement of a crayfish outside their natural ranges: (i) natural ones, such as

48

active dispersal, (ii) accidental, such as escape from holding facilities, or (iii) deliberate, by humans (Barbaresi et al., 2003). Population genetics may be used to discover, which method is at the source of the distribution of the invasive species, and identifying the source population could be helpful in understanding the invasion and implementing control efforts (Barbaresi et al., 2003). The genetic diversity of all the sites may point to bait dumping as a source of many of these crayfish populations, most prominently the Institute Pond site. The Institute Pond site was started and populated, from our data, by one female. This female may have been the only remaining female in the dumped bait, or a pregnant crayfish that was dumped into the pond. The average genetic diversity of the other sites suggests, if a bait dumping occurred, that a larger number of crayfish were dumped. This could have happened over time as fisherman dumped their bait into the same waters every time they had extra bait. Depending on the source of the bait crayfish, this would allow for some means of gene flow, which results in average genetic diversity. However, if the species is found not to be invasive, which the data could also suggest due to average genetic diversities, then the source of gene flow could be investigated.

The fact of figuring out whether or not the species is invasive must first be investigated. Once that is known, as stated above it may be easier to understand how and begin to take a course of action for or against the species, *Orconectes virilis*, in Massachusetts depending on the results.

WORKS CITED

Alfaro ME, Zoller S, Lutzoni F (2003) Bayes or bootstrap? a simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20**, 255-266.

Avise JC (2000) *Phylogeography* Harvard University Press.

Barbaresi S, Fani R, Gherardi F, Mengoni A, Souty-Grosset C (2003) Genetic variability in European populations of an invasive American crayfish: preliminary results. *Biological Invasions* **5**, 269-274.

Beebe TJC, Rowe G (2004) *An introduction to molecular ecology* Oxford University Press, New York, USA.

Beerli P (2005) Comparison of Bayesian and maximum likelihood inference of population genetic parameters. *Bioinformatics* **22**, 341-345.

Bowles I (2007) French/Quinebaug rivers watershed. Commonwealth of Massachusetts, Boston, MA.

http://www.mass.gov/envir/water/frenchQuinebaug/frenchQuinebaug.htm.

Bubb DH, Thom TJ, Lucas MC (2004) Movement and dispersal of the invasive signal crayfish *Pacifastacus leniusculus* in upland rivers. *Freshwater Biology* **49**, 357-368.

Buhay JE, Crandall KA (2005) Subterranean phylogeography of freshwater crayfishes shows extensive gene flow and surprisingly large population sizes. *Molecular Ecology* **14**, 4259-4273.

Bussack CA (1988) Electrophoretic variation in the red swamp (*Procambarus clarkia*)
and white river crayfish (*P. acytys*). *Aquaculture* **69**, 211-226.

Byron CJ, Wilson KA (2001) Rusty crayfish (*Orconectes rusticus*) movement within and
between habitats in trout lake, Vilas County, Wisconsin. *Journal of the North
American Benthological Society* **20**, 606-614.

Castro LR, Dowton M (2007) Mitochondrial genomes in the Hymenoptera and their
utility as phylogenetic markers. *Systematic Entomology* **32**, 60-69.

Cavalli-Sforza LL (2005) The Human Genome Diversity Project: past, present and
future. *Natural Reviews Genetics* **6**, 333-340.

Clewley JP (1998) A user's guide to producing and interpreting tree diagrams in
taxonomy and phylogenetics: part I. introduction and naming of parts. .
Communicable Disease and Public Health **1**, 64-66.

Costa FO, Neuparth T, Theodorakis CW, Costa MH, Shugart LR (2004) RAPD analysis
of southern populations of *Gammarus locusta*: comparison with allozyme data
and ecological inferences. *Marine Ecology Progress Series* **277**, 197-207.

Crandall KA (1998) Conservation phylogenetics of Ozark crayfishes: assinging priorities
for aquatic habitat protection. *Biological Conservation* **84**, 107-117.

Crocker DW, Barr DW (1968) *The handbook of crayfish of Ontario* University of
Toronto Press.

Darwin C (1859) *On the origin of species by means of natural selection or the
preservation of favored races in the struggle for life.* John Murray, London,
England.

Donoghue MJ, Kadereit JW (1992) Walter Zimmerman and the growth of phylogenetic theory. *Systematic Biology* **41**, 74-85.

Effron B, Halloran E, Holmes S (1996) Bootstrap confidence levels for phylogenetic trees. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 13429-13434.

Emerson BC, Hewitt GM (2005) Phylogeography. *Current Biology* **15**, 367.

Excoffier L, Laval G, Schneider S (2005) Arlequin version 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47-50.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Fetzner Jr. JW, Sheehan RJ, Seeb LW (1997) Genetic implications of broodstock selection for crayfish aquaculture in the Midwestern United States. *Aquaculture* **154**, 39-55.

Finlay JB, Buhay JE, Crandall KA (2006) Surface to subsurface freshwater connections: phylogeographic and habitat analyses of *Cambarus tenebrosus*, a facultative cave-dwelling crayfish. *Animal Conservation* **9**, 375-387.

Gilbertson L (1999) *Zoology lab manual* McGraw Hill Companies, New York.

Gottardi E, Anderson E (2006) *Population genetics of invasive crayfish (Orconectes virilis) in massachusetts*, Worcester Polytechnic Institute.

Grandjean F, Souty-Grosset C (2000) Mitochondrial DNA variation and population
genetic structure of the white-clawed crayfish, *Austropotamobius pallipes
pallipes*. *Conservation Genetics* **1**, 300-319.

Hall TA (1999) Bioedit: a user-friendly biological sequence alignment editor and analysis
program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**, 95-98.

Halliburton R (2004) *Introduction to population genetics* Pearson Prentic Hall, Upper
Saddle River, NJ.

Hare MP (2001) Prospects for nuclear gene phylogeography. *Trends in Ecology and
Evolution* **16**, 700-706.

Hazlett BA, Rittschof D, Ameyaw-Akumfi C (1979) Factors affecting the daily
movement of the crayfish *Orconcectes virilus*. . *Crustaceana* **5**, 121-130.

Hobbs Jr. HH (1989) An illustrated checklist of the American crayfishes. *Smithsonian
Contributions to Zoology* **480**, 236.

Horwitz P (1990) The conservation status of Australian freshwater crustacean with a
provisional list of threatened species, habitats, and potentially threatening
processes. Australian National Parks and Wildlife Services, Canberra, Australia.

Horwitz P (1994) Distribution and conservation of the Tasmanian giant freshwater
lobster *Astacopsis gouldi*. *Biological Conservation* **69**, 199-206.

Horwitz P (1995) The conservation status of Australian freshwater crayfish: review and
update. *Freshwater Crayfish* **10**, 70-80.

Huelsenbeck J, Rannala B (1997) Phylogenetic methods come of age: testing hypothesis
in an evolutionary context. *Science* **276**, 227-232.

Huelsenbeck J, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 745-755.

Hunter JV (1994) *Freshwater crayfish aquaculture in North America, Europe, and Australia: families Astacidae, Cambaridae and Parastacidae* Haworth Press, New York.

Iizuika MA (1989) A population genetic model for sequence evolution under mulitple types of mutation. *Genetic Research* **54**, 231-237.

Interior UDo (2007) French River Basin- surface water. http://ri.water.usgs.gov/basins/french.htm (ed. Center UM-RIWS).

Kerr M (1990) The Blackstone River. In: *Rhode Island Sea Grant Fact Sheet*. *http://seagrant.gso.uri.edu/factsheets/blackstone_river.html*.

Knowles LL, Maddison WP (2002) Statistical phylogeography. *Molecular Ecology* **11**, 2623-2635.

Kolbe JJ, Glor RE, Schettino LR, Lara AD, Larson A (2004) Genetic variation increases during biological invasion by a cuban lizard. *Nature* **431**, 177-181.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML. PN (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyes. *Proceedings of the Natural Academy of Sciences of the United States of America* **82**, 6955-6959.

Largiader CR, Herger F, Lortscher M, Scholl A (2000) Assessment of natural and artificial propagation of the white-clawed crayfish (Austropotamobius pallipes species complex) in the Alpine region with nuclear and mitochondrial markers. *Molecular Ecology* **9**, 25-37.

Lee LG, Connell CR, Woo SL, Cheng R, McArdle BF FW, Halloran ND, Wilson RK
(1992) DNA sequencing with dye-labeled terminators and T7 DNA polymerase:
effect of dyes and dNTPs on incorporation of dye-terminators and probability
analysis of termination fragments. . *Nucleic Acids Research* **20**, 2471-2483.

Letunic I, Bork P (2006) Interactive Tree of Life (iTOL): an online tool for phylogenetic
tree display and annotation. *Bioinformatics* **23**, 127-128.

Li S (1996) *Phylogenetic tree construction using Markov Chain Monte Carlo* Ph. D.
Dissertation, Ohio State University.

Li WH (1997) *Molecular evolution* Sinauer Associates, Sunderland, Massachusetts.

Lycott (1989) *Diagnostic/feasibility study for the management of Indian Lake, Worcester,*

MA. Lycott Environmental Research, Inc., Southbridge, MA.

Mathews L (2006) Cryptic biodiversity and phylogeographical patterns in a snapping
shrimp species complex. *Molecular Ecology* **15**, 4049-4063.

Nei M (1987) *Molecular evolutionary genetics* Columbia University Press, New York,
NY.

Nylander JAA (2004) MrModeltest (version 2.2), Program distributed by the author.
Evolutionary Biology Centre, Uppsala University.

Purugganan M, Gibson G (2003) Merging ecology, molecular evolution, and functional
genetics. *Molecular Ecology* **12**, 1109-1112.

Reynolds JD (2002) *Biology of freshwater crayfish* University of Nottingham,
Nottingham, UK.

Richardson AMM, Doran N, Hansen B (1999) The conservation status of Tasmanian freshwater crayfish. *Freshwater Crayfish* **12**, 863-877.

Rieseberg LH (2000) Phylogeography. book review of phylogeography; the history and formation of species, by John Avise. *Trends in Ecology and Evolution* **15**, 384-385.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R HG, Mullis KB, Erlich HA (1988) Primer-detected enzymatic amplification of DNA with thermostable DNA polymerase. *Science* **239**, 487.

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the Natural Academy of Sciences of the United States of America* **74**, 5463-5467.

Sbordoni V (1982) Advances in speciation of cave animals. In: *In Mechanisms of Speciation* (ed. Barigoozi C), pp. 219-240. Alan Liss, Inc., New York.

Stockard CR (1908) The influence of regeneration on moulting in *Crustacea*. *The American Naturalist* **42**, 140-142.

Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437-460.

Taylor CA, Warren Jr ML, Fitzpatrick JF, Hobbs III HH, Jezerinac RF, Pflieger, W. L., and Robinson H. W. (1996) Conservation status of crayfishes of the United States and Canada. *Fisheries* **21**, 25-38.

Tsutsui ND, Suarez AV, Holway DA, Case TJ (2000) Reduced genetic variation and the success of an invasive species. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 5948-5953.

Voet D, Voet J, Pratt C (2006) *Fundamentals of biochemistry: life at the molecular level.* John Wiley and Sons Inc, Philadelphia, PA USA.

Wandle Jr. SW, Phipps A (1984) Gazetteer of hydrologic characteristics of streams in Massachusetts-coastal river basins of the North Shore and Massachusetts Bay. USGS Water-Resources Investigations Report 84-4281.

Wu C (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology* **14**, 851-865.

Zimmerman W (1934) Research on phylogeny of species and of single characters. *The American Naturalist* **68**, 381-384.