Project 0915

INVESTIGATION OF INTROGRESSION IN THE BLACKSTONE RIVER VALLEY: LOCATING AFLP MARKERS

A Major Qualifying Project Report:

Submitted to the Faculty of

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Bachelor of Science

Submitted By:

Margaret Marie Allard

April 29, 2010

Approved:

_____

Professor Michael A. Buckholt

_____

Professor Lauren Mathews

Abstract:

Introgression may be occurring between the species *Orconectes quinebaugensis* and a related invasive species, *Orconectes virilis*, in the Blackstone River Valley Area. While genetic analysis has already been used to investigate this event, more robust conclusions can be drawn if more data is generated. To fill this need, genetic differences between the two species were sought by performing AFLP analysis on non-hybrid members of each species. Eight members of non-hybrid *Orconectes virilis* were tested against four non-hybrid *Orconectes quinebaugensis*. Among these, twelve fragments whose occurrence correlates with species identity were located. The author recommends that these twelve markers be verified by screening a larger pool of organisms. Once confirmed, they may be used to study genetic exchange between the two species in the Blackstone River Valley Area.

Acknowledgments

I would like to thank Professor Michael A. Buckholt and professor Lauren Mathews for their guidance on this project. Their comments, loaned materials, and instruction in using laboratory equipment throughout the project have been greatly appreciated.

Table of Contents

Abstract: 2

Acknowledgments 3

Introduction 5

Materials and Methods 25

Samples & DNA Extraction 25

Results and Discussion 31

Works cited 38

Appendix: 41

A- AFLP: Basic Principles 41

B- Additional Data 47

Introduction

Previous studies indicate that introgression may be occurring between the two freshwater crayfish species, O*rconectes virilis* and O*rconectes quinebaugensis,* within the Blackstone River Valley watershed (BRV; Mathews, personal communication). If this is true, yet-undiscovered genetic markers may exist that correlate with species designation at collection sites where the species are still isolated from each other, but that occur in both species types in collection sites where the two species coexist. Studying an introgression event like this one is important, because it may have dire effects for biodiversity in affected areas.

In this study, I searched for quantifiable genetic differences between *Orconectes virilis* and *Orconectes quinebaugensis*. For this task, I used amplified fragment length polymorphisms (AFLPs), one tool of many in the field of molecular ecology. In this section, I provide some background information relevant to the phylogeography and biology of the two species and the methods used to study them.

Phylogreography is a fairly young discipline which addresses and relates such diverse fields as biology, paleontology, and historical geography. The word was coined in 1987 by John Avise (Avise et al, 1987), who says in his recent book that this coining was for utilitarian purposes. Acording to him, earlier researches had to use cumbersome phrases to communicate a straightforward observation, which was that branches in intraspecific gene trees often displayed geographic patterns. The word phylogeography came to mean the field of study concerned with the processes governing the relationship between genetic lineages, geographic location, and time (Avise 2000).

Unlike other fields, which emphasize the role of natural selection upon biodiversity, phylogeography highlights historical events, such as geographic isolation of groups from the same species. In cases like these, the absence of the homogenizing influence of gene flow between the isolated groups allows for evolutionary divergence through selectively neutral genes. A primary cause of disparity between the descendents of the isolated groups would be random occurrences and imperfections in the process of DNA replication. Phylogeographers try

to map the effects of such historical events upon contemporary geographic distributions of gene-based traits. Oftentimes, it is possible to infer the occurrence of ancient events for which no record exists through observations of this distribution (Avise 2000).

One natural event which falls well within phylogeography's domain is introgression, which implies hybridization. Introgression and hybridization are rare ecological events that confound the taxonomical process (Freeland 2005). A hybrid is an organism produced by the interbreeding of two groups that are from taxonomically distinct species. Hybridization is the production of offspring between male and female members of different species, while introgression is the transfer of genetic material from one species to another following hybridization (Trevor and Rowe, 2008). This occurrence renders useless the Biological Species Concept (May, 1942), which has traditionally defined different species as those groups that cannot interbreed to produce fertile offspring (Trevor and Rowe, 2008).

Hybrids are generally rare, but some localities are an exception to this rule. A hybrid zone is an area where species are interbreeding and producing hybrids at an unusually high rate. For example, in cases of primary contact, a single species gradually spreads over a contiguous area. One segment of the population suddenly spreads into an adjacent but very dissimilar area, with dissimilar environmental pressures. This population begins to adapt and change, while maintaining contact with the more static group that remains in the original environment. Contact occurs in the hybrid zone which forms between the two areas. Alternatively, in secondary contact, two closely related species have been geographically isolated (Freeland 2005). Their separation might be caused by something as monumental as the movement of glaciers across a landscape, or as incidental as the sudden relocation of a small group by extreme weather (a hypothesized mechanism for populating young islands) (Neal 2004). Once isolated, the population begins to diverge, either due to genetic drift or a combination of genetic drift with dissimilar environmental pressures. At some point, the groups regain contact, which is definitely secondary contact. Their reconnection could be due to natural or artificial changes. A natural change might be one stream diverting into another, while an artificial

change would be any human mediated event (Freeland, 2005). This second type of contact is occurring in the case under study (Mathews et al, 2008).

Hybridization and introgression affects more than systematics. Published reviews have speculated on numerous results of hybridization, from an increase of biodiversity in a location to the elimination of one species by assimilation (Wolf et al, 2001, Rhymer & Simberloff 1996, Ayers, et al, 1999). In some cases, phenotypic variation in a hybrid population will be greater than the combined variation in both parent species. This is known as transgressive segregation, and has been observed in plants (Sweigart and Willis, 2003). If some of the variation produced in transgressive segregation is adaptive to the environment, it may lead to speciation.

One example of increased genetic diversity due to hybridization is the famous case of Darwin's ground finches. These finches undergo cyclic declines and increases in population. During extremes such as drought or famine, the population bottlenecks to a fraction of the preceding generation. When the environment becomes more tenable, the population rapidly increases to previous levels. One would expect to observe depleted genetic diversity in a population which is regularly reduced to a few members, but this is not the case. Hybridization with five other species of finch that populate the Galapagos seems to provide the ground finches with a variety of genetic material (Freeland and Boad, 1999).

On the other hand, According to Wolf et al. (2001), natural hybridization poses the risk of extinction to numerous plant and animal species. Ayers et al. (1999) characterized one extensive example: the invasion of the American west coast by the smooth cord grass *Spartina alterniflora*. Genetic analysis confirms that the invasive *Spartina alterniflora* has hybridized with native competitor *Spartina foliosa;* to the extent that some parts of San Francisco Bay contain only *S. alterniflora* and hybrids. The fact that this transformation took less than forty years shows how rapidly hybridization can affect the biodiversity of a location.

Another example of the elimination of a species through introgression was characterized on a smaller scale by Perry et al (2000). In this case, an invasive species, *Orconectes rusticus*, were introduced into a Wisconsin lake, the native habitat of a closely related species, *Orconectes*

propinquus. Researchers hypothesized that the superior size of the *O. rusticus* F1 males would allow them to outcompete native *O. propinquus* males for *O. propinquus* females. Because it was believed that most of the hybrids produced by such a mating would be either sterile or suffer from some other genetic maladaptation, researchers expected to see a rapid decline in the *O. propinquus* population due to decrease reproductive output by *O. propinquus*. Instead, they observed something very different. The hybrids produced between the two species were not only fertile, but also possessed phenotypes that allowed them to outcompete both parent species for limiting resources. While the number of pure *O. propinquus* organisms decreased, and is still projected to disappear, nuclear genes of the species were perpetuated through the hybrid population. Meanwhile, the spread of *O. rusticus* genes in the habitat proceeded at a much faster rate than expected (Perry et al, 2000). This example is particularly relevant to the current study, as it occurs within the same genus as the two species of interest.

Because of human interference, freshwater systems, such as those where the species of interest in this study reside, may be particularly vulnerable to a hybridization event. Jenkins (2003), when comparing the recorded biodiversity loss between terrestrial, marine, and freshwater environments, notes a significantly greater decline in freshwater environments. Contact between native and non-native species may happen at a higher rate in freshwater areas due to human-mediated events like the release of ballast water (Kolar & Lodge 2000; Ricciardi & MacIsaac 2000; Holeck et al. 2004). In cases like these, water that has been loaded onto ships for buoyancy purposes in one location is ejected into the waters of another location. Any whole organisms suspended in the ballast water, such as live crayfish, are thus relocated. Kolar and Lodge mention a case in 1991, wherein one hundred thousand Peruvians were killed by a strain of *Vibrio cholera* that had been released in their drinking water. They speculate that completely halting the spread of species by ballast water may be impossible. However, they note that taking measures to reduce the number of individuals relocated in this manner will probably reduce the rate at which populations become firmly established in new areas. Additionally, according to Padilla & Williams (2004), the release of non-native aquarium animals into freshwater streams is a major cause of species invasion.

In studying events like these, the phylogeographers must use the tools of molecular ecology, which generate the data that is fodder for phylogeographical analysis. In fact, the two disciplines are so intertwined that one cannot understand phylogeography without first understanding molecular ecology.

General ecology, from which molecular ecology sprang, is a branch of biology that is concerned with how organisms interact with their environment and with each other. For example: How are ecosystems maintained? What is the natural habitat for the species under study? How closely is it related to other species, and what is its evolutionary history? Before the 1960's, scientists had to rely on phenotypic data to answer questions like these.

Phenotypic data is any observation related to the behavior, morphology and physiology of an organism (Freeland, 2005). While it can be informative, this type of data is limited, because it is far removed from the genetic code. One way to visualize this problem is presented below. This chart, taken from Swallow and Garland (2009), shows the influence of natural selection on various facets of an organism. The order of these characteristics reflects their relative susceptibility to environmental influences.

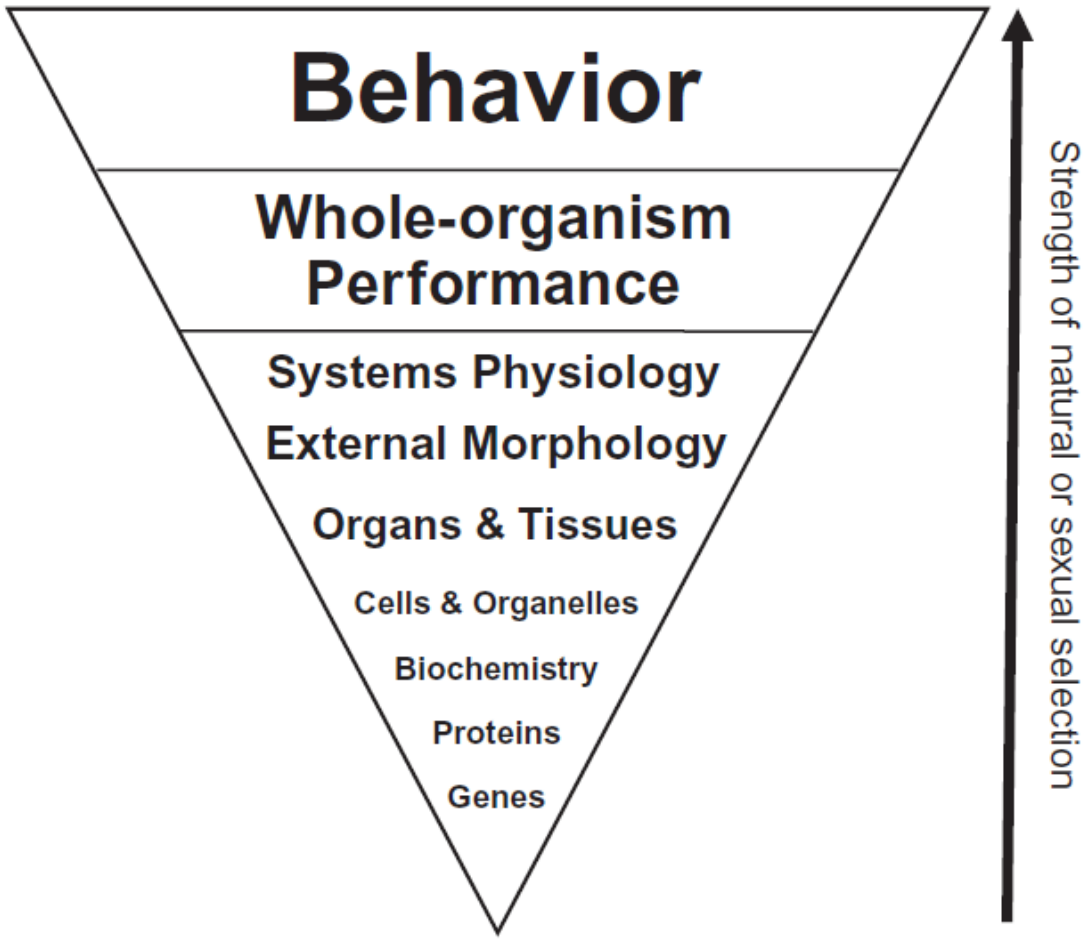

Figure 1: Strength of Environmental Influences (from Swallow and Garland, 2009)

Phenotypic traits (in this figure, all those qualities listed above "Genes") can be the result an organism's genetic heritage, but this link can be confounded by environmental influences. For example, experiments have shown that genetically dull rats, when reared in a stimulating environment, can perform as well as genetically bright rats at learning tasks (Cooper and Zubeck, 1958). Additionally, the challenges presented by similar environments can result in convergent evolution of similar phenotypic traits in genetically dissimilar and historically remote species. For example, according to Emery & Clayton (2004), crows and apes are undergoing convergent evolution in cognition. Each line has separately developed a similar 'tool kit' for cognition, including causal reasoning, flexibility, imagination, and prospection.

Such resemblances have nothing to do with how closely related the two species are, and can be misleading. Thus, phenotypic data is limited.

Molecular ecology side-steps the problems of ambiguity in phenotypic data by 1) being able to detect differences in the genes responsible for causing phenotypically similar traits and 2) by examining selectively neutral traits, as referred to above. Selectively neutral traits abound in the genetic code. Sequence information in noncoding regions of DNA can be vastly different between two organisms without affecting their phenotypic similarity, and thus their relative fitness. Unlike phenotypes that add to or detract from an organism's fitness, which tend to propagate or vanish due to their effects upon an organism's fitness, selectively-neutral traits tend to occur due to imperfections in the copying mechanism for DNA, described in more detail below. Once generated, selectively neutral traits may propagate of vanish through random chance (Trevor and Rowe, 2008). Thus, conclusions about the evolutionary history and relatedness of species bases on such data should be considered tentative, while more direct observations of genes are more reliable.

Molecular ecology emerged when new ways of generating and analyzing data were developed. These techniques made genetic data more directly accessible. The event that some consider the birth of molecular ecology is a study by Harris (1966), and concurrently by Lewontin and Hubby (1966). These scientists developed a method to quantify genetic variation by identifying structural differences in proteins, the direct products of genes (Freeland, 2005).

Later, techniques using molecular markers were developed that analyze genes even more directly, by detecting differences in DNA sequences themselves. Two frequently used media in this category are SNPs and microsatellites (Trevor and Rowe, 2008). DNA is transmitted from parent to offspring in a fairly predictable manner. Because of this, differences in the genetic code from one individual to another can allow one to infer genetic relationships through millions of generations. These differences come in several distinctive forms, each caused by a different inaccuracy in the copying process of DNA. These forms are the types of molecular markers (Freeland, 2005).

One commonly used marker is the microsatellite. Microsatellites, also known as simple sequence repeats (SSR), short tandem repeats (STR), or variable number tandem repeats (VNTR), are short stretches of DNA found in organisms that consist of tandem repeats, usually about 1-6 base pairs long. Microsatellites tend to occur in chloroplasts and mitochondria, but they can also be found in nuclear DNA. When they occur near or in coding regions, microsatellites can have detrimental effects. For example, the (CGG) 30 repeat found in humans suffering from Fragile X Mental Retardation Syndrome is responsible for their dysfunction (Jara et al, 1998).

Microsatellites tend to exhibit a high rate of mutation (10^-2 to 10^-6 nucleotides per locus per generation (Sia et al, 2000)) compared to other gene regions. It is believed that this high rate is due to the foibles of meiosis and DNA replication: either unequal crossing over during meiosis, or strand slipping during replication (Oliveira et al, 2006). In crossing over, the homologous regions present in long repeats can allow hairpins (single-stranded kinks in DNA) to form, causing chromosome segments of unequal lengths to exchange. In this scenario, one chromosome will receive a larger microsatellite segment, while the other chromosome will decrease the length of its microsatellite region (Oliveira et al, 2006). This process is pictured in the figure below.

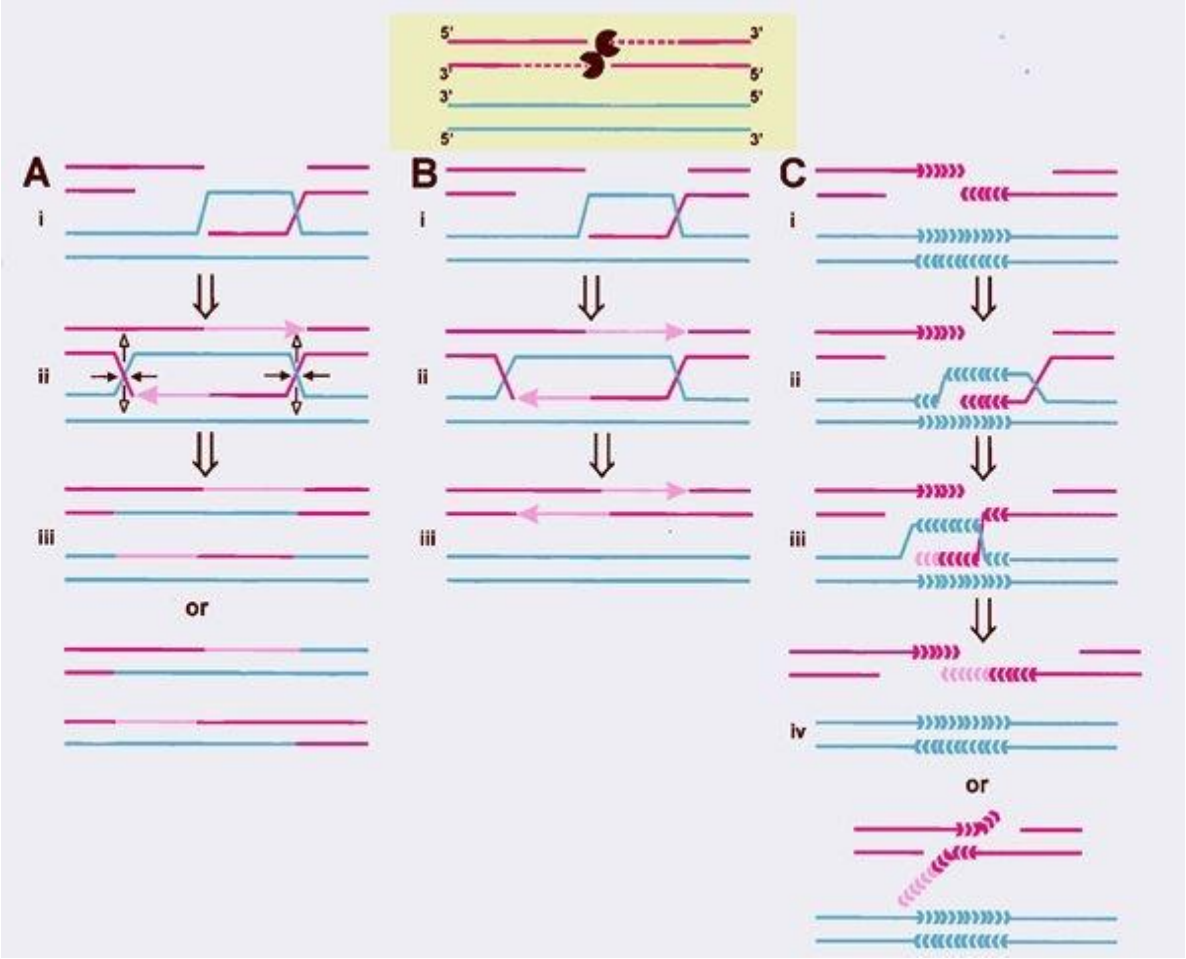

Figure 2. Unequal Crossing Over due to the Non-Sequential Annealing of Homologous Regions (from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=eurekah&part=A13297&rendertype=figure&id=A13312)

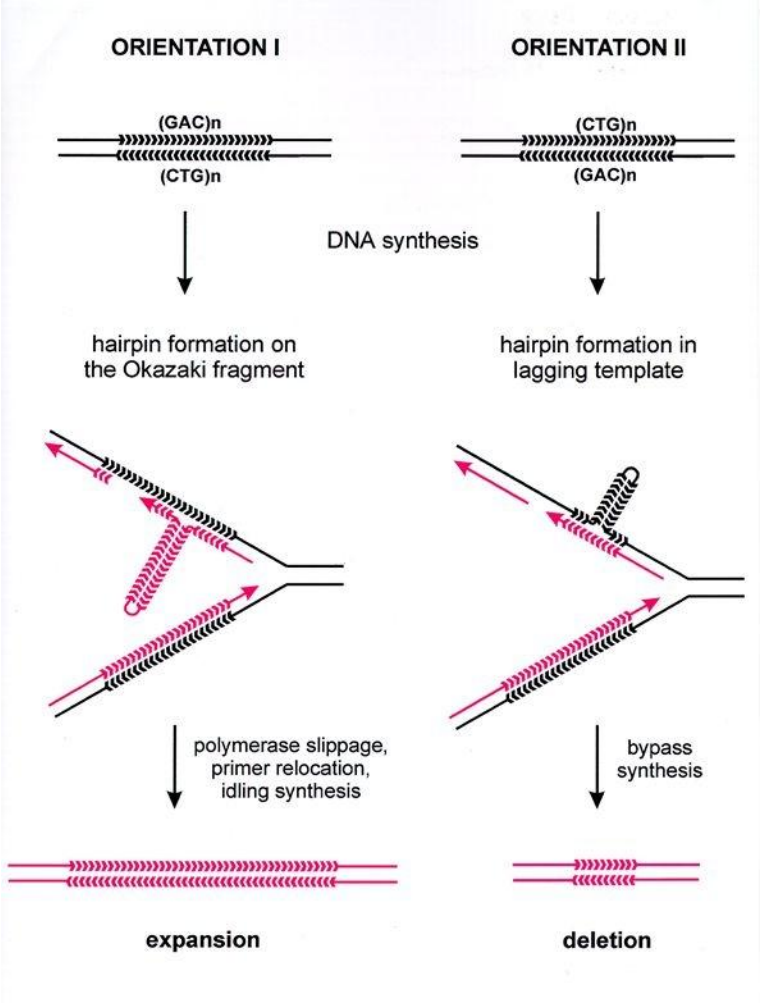

Figure 3. Strand Slippage Increases or Decreases the Number of Repeats in a Microsatellite (from http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=eurekah&part=A13297).

Strand-slippage, pictured above in Figure 3, is also compounded by the vast homologous regions in microsatellites. During DNA replication, as DNA polymerase travels along a single parent strand of DNA while assembling the complementary strand, it sometimes "slips" off the parent strand. If its trailing, newly formed DNA sequence contains homologous regions, it may fold over on itself, causing a single-stranded hairpin to form. The same folding potential applies to the parent strand. This bunching causes the polymerase, when it is reattached to the parent strand, to fail to read the hairpinned regions of the parent strand, or to begin reading several frames earlier than where it left off. This causes a discrepancy between the length of the parent and complementary DNA strands (Oliveira et al, 2006).

Microsatellites, due to their high rates of mutations, are a hotbed for genetic variation between even closely-related individuals. However, size homoplasy, caused by a microsatellite region's ability to both increase and decrease in length by chance, can create markers with dissimilar histories that are identical in length and sequence. For this reason, microsatellites are better suited to discern events that occurred in the recent, rather than the distant, past (Oliveria et al, 2006).

Another commonly used molecular marker is the SNP, or single nucleotide polymorphism. This term refers to single base pair positions in a DNA sequence that may differ between individuals. Usually SNPs are bi-allelic, meaning that for one site there are only two common possibilities (for example, either adenine or cytosine). SNPs are extremely common relative to other forms of variability in the genetic code. According to Collins et al. (1988), SNPs account for 90% of the variation within the human genome. A single nucleotide can mutate for a number of reasons, from oxidative damage to radiation (Deikman et al, 2009). They characterize a particular example of SNPs that, like microsatellites, are caused by hairpining. See Figure 2 below.

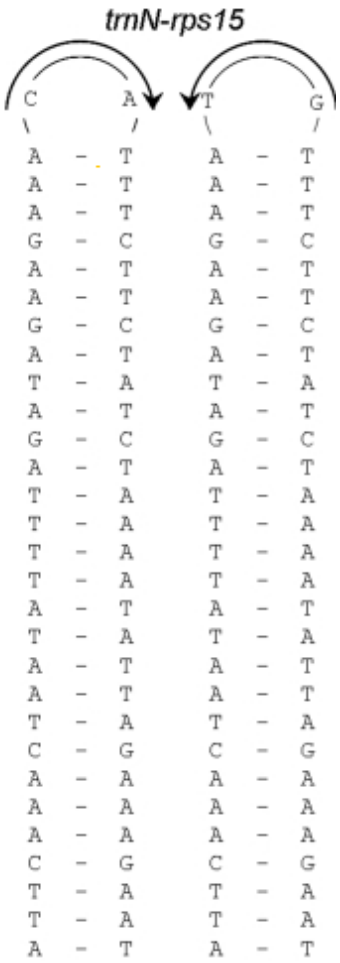

Figure 4. Hairpin Loop found in the Chloroplast Genome of *Lolium Perenne* (Deikmann, et all, 2009)

As figure two illustrates, homologous regions in DNA can result in the inversion of nucleotides, and thus the appearance of multiple SNPs in the genetic code of an organism. Even if these multiple nucleotide shifts are not technically "single" nucleotide polymorphisms, Deikmann et al (2009) still apply this term, because the variance resembles an SNP when SNP detection methods are used.

Unlike microsatellites, the mutation rate of SNPs appears to be very low, on the order of 10^{-8} mutations per allele transfer from parent to offspring (Brumfield et al, 2003). This mean that they are stable, tending not to occur very often and tending to not be reversed in the descendents of the original mutant. Therefore, SNPs are most useful for identifying long branches or key phylogenetic positions. In other words, SNPs are most useful when addressing

questions about the distant past (Kiem et al, 2004). Thus, the two approaches complement each other.

In order to use either microsatellites or SNPs in a comparative study, they must be first identified within the genome. SNPs can be located by a number of different tools, including base specific primer extension, and the more recent high-density variant detection array technology. The former is depicted below:

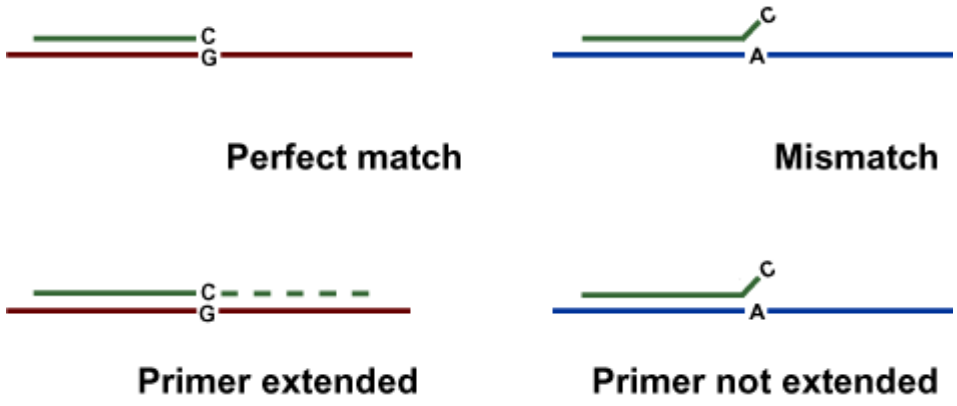

Figure 5. Allele-specfic Primer Extensions (http://las.perkinelmer.com/content/snps/protocol.asp): in this figure,

Base specific primer extension is a very sensitive but inefficient process (Dong, 2001). Meanwhile, high density variant detection arrays are much more efficient, but much more prone to error. In this process, genomic DNA is divided by restriction enzyme digestion and gel-based fragment size separation. The separated fragments are ligated to a common adaptor, and amplified with one primer in a single PCR (see appendix for explanation of PRC) reaction. Variations are screened by their ability to hybridize with complementary sequences on an array.

Microsatellites also require a significant initial discovery effort. Using PCR, the investigator makes many copies of the variable region of interest, while leaving the amount of genomic DNA unchanged. DNA gel electrophoresis is then used to separate amplified microsatellites by size (Freeland, 2005). As mentioned earlier, however, size is not an infallible record of their history.

These are just two examples of molecular marker techniques. Both have been used previously to study the subjects involved in this experiment. The method used in this particular study was Amplified Length Polymorphism analysis, which is related to restriction fragment length polymorphism (RFLP) and random amplification of polymorphic DNA (RAPD), as described below (Benchs and Akesson, 1996).

Restriction fragment length polymorphism refers to the laboratory technique that differentiates between fragments of different lengths generated by restriction digestion (see appendix for a discussion of restriction digestion). The fragments are separated by size via gel electrophoresis. The differences in fragment length between samples from different organisms reflect differences in position of restriction sites that are recognized by the digestion enzymes used in the reaction from one organism to another. For example, a fragment of length X base pairs (bp) indicates that the DNA used in this RFLP had at least two restriction sites exactly X base pairs apart. The destruction of one of those restriction sites through a mutation would prevent digestion at that site, increasing the length of the fragment to X + Y, Y being the number of base pairs between the eliminated cut site and the nearest cut site along the DNA strand. Thus, the fragments of length X and of length Y would disappear from gel array, and the fragment of length X+Y would appear. Other events that could change the length of fragments in RFLP are insertion and deletion mutations, which, when occurring between two restriction sites on fragment A, would increase or decrease the length of fragment B, respectively. Because mutations in the genome affect the output of RFLP analysis, and because so many fragments are generated per run, this type of analysis can be used to screen for the presence and absence of many mutations at once. However, it is not particularly useful in discerning the type of mutations it detects, unless coupled with sequencing technology (Jeffrys et al, 1986).

Random amplification of polymorphic DNA, on the other hand, is a PCR technique that uses short (about 10 bp) primers of an arbitrary sequence to amplify a small fraction of a DNA sample. Only portions of the DNA strand which fall between areas complementary to the chosen primers within a certain distance of each other are amplified. Once generated, the amplified fragments, the markers, can be isolated and sequenced. Using the now-available

sequence information, longer and more specific primers can be designed for particular markers. The technique can distinguish the presence and absence of a small number of genetic variabilities, but again, information about the regions between the sites of interest cannot be understood using this tool alone. Other limitations include the inability to discern whether a dominant or recessive fragment is being amplified, and the ambiguity of results in cases of nearly complementary primers. In the case of near complementary, the output may be either an absence of the fragment of interest or a reduced amount of the product (Lynch and Miligan, 1994).

In addition to background related to phylogeography, molecular ecology, and molecular marker systems, it is also vital to the understanding of this project to be familiar with certain aspects of the two species under study, *Orconectes virilis* and *Orconectes quinebaugensis*.

O. quinebaugensis is morphologically very similar to *O. virilis*. The major distinguishing external feature between the two is the curvature and lengths of the rami of the first pleopods of adult males, as well as the shape of the dactyl of the chelipeds (Mathews and Warren, 2008).This paper was the work in which the discovery of the creature as a unique species was announced and supported.

Members of the *Orconectes* genus are found in freshwater environments such as lakes, streams, and rivers. Typically, they are found in rocky-bottomed bodies of water, but have also been observed in silty or stagnant lakes and marshes. They shelter in burrows constructed under submerged rocks, logs, or aquatic plants (Hamr, 2002). As referred to above, freshwater systems such as these may be particularly vulnerable to hybridization via secondary contact, due to human-mediated events.O. virilis is present in a vast native range, but the exact site of the species' origin is unknown. According to Hamr (2002), this natural range extends across a large portion of North America. In a small portion of this range, this widespread species is being displaced by its invasive cousins, *O. limosus* and *O. rusticus*.

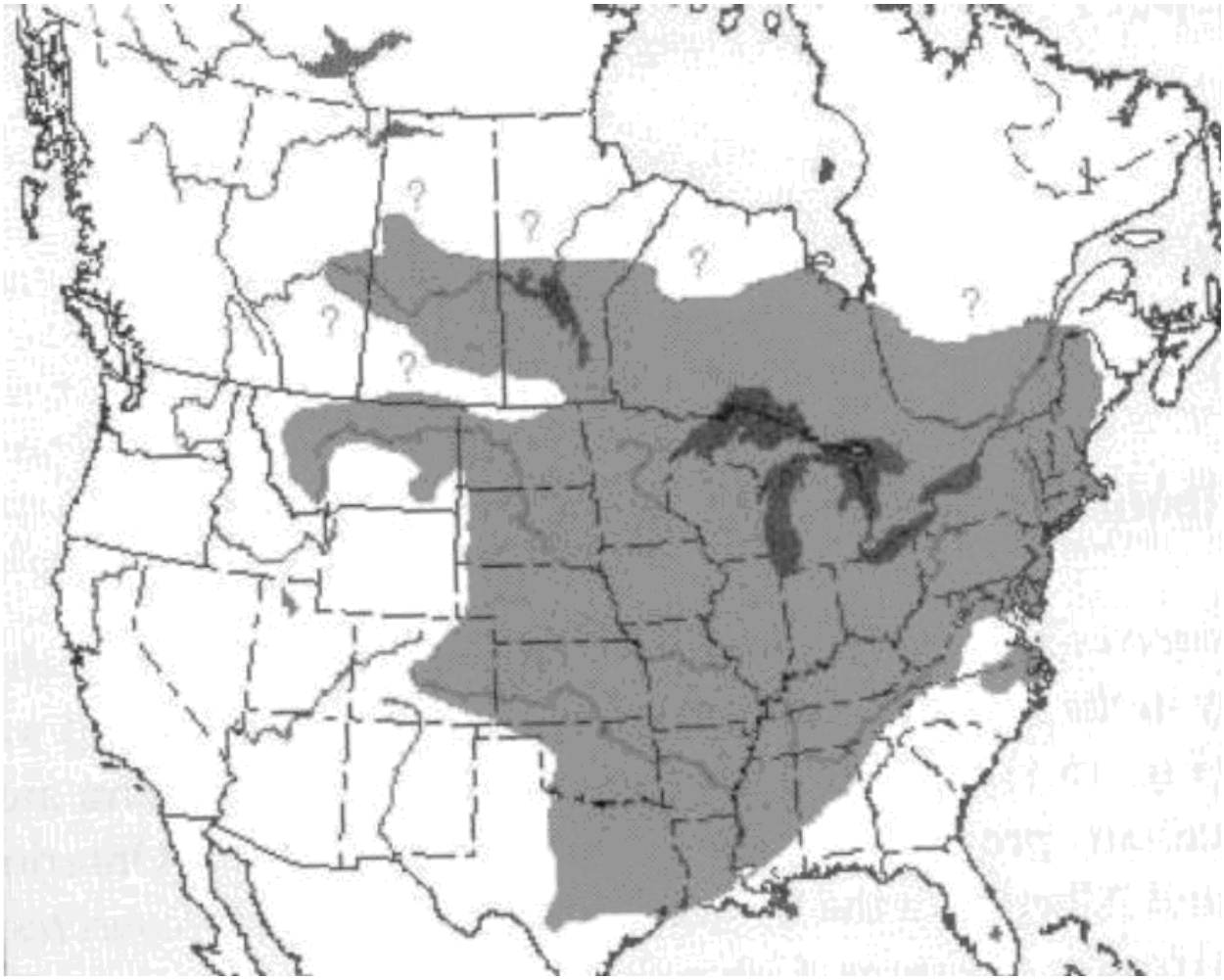

Figure 6. Native Range of *O. virilis* (Harm, 2002)

Outside of its native range, *O. virilis* have been introduced in a number of places across New England, the southern United States, Mexico, and even in France and Sweden (Harm, 2002). The presence of *O. virilis* in Massachusetts, while extensive, is considered invasive, as indicated by the Global Invasive Species Database (http://www.issg.org/database/).

Meanwhile*, O. quinebaugensis* has been only recently characterized (Mathews & Warren, 2008). Thus far, *Orconectes quinebaugensis* has been found in only small streams in Massachusetts with rocky or gravel substrate, including the Quinebaug River in southern New England, for which it was named. *O. quinebaugensis* is hypothesized to be native to this area, though more data are needed to better understand its evolutionary origins (Mathews and Warren, 2008).

Mathews et al. (2008) confirmed a close relationship between the two species. Using the COI and 16s rRNA mitochondrial genes, Mathews (2008) estimated that *O. virilis* and *O. quinebaugensis* had diverged from a common ancestor approximately 1.8 to 2 million years ago, assuming clock-like mutation of those genes. McMurrough and Saltzman (2009) hypothesized that the species was split physically into two groups, one in the Midwest, and the other in New England. A likely cause of this hypothesized separation is the movement of glaciers during Pleistocene era.

McMurrough and Saltzman (2009) assert that during the two million years that the groups were kept apart, genetic drift allowed for a number of slight genetic and phenotypic differences to emerge between them. For example, the male gonopods in each species are slightly different, and potentially not well adapted to mate with females of the other species. While the groups were physically separated, inability to mate with the other group would have been a selectively neutral trait, and so the trait was able to be propagated (Mathews et al, 2008).

This project is a further investigation of an apparent example of hybridization occurring in a freshwater environment, between the crayfish species *O. virilis* and *O. quinebaugensis* in the Blackstone River Valley area (see Figure 6). Unpublished data (Mathews, personal communication) examining both morphology and nuclear DNA, indicates that hybridization is occurring between these two species. Additional data from the nuclear genome in the form of AFLPs would allow for inferences that are more robust.

Previous research teams (Mathews personal communication, McMurough and Saltzman, 2009, Becker et al, 2009) examined this population in the following manner. First, experimenters made the assumption that the location R1 contained a pure or nearly pure population of *O. Virilis*, and that location M2 was a pure population of *O. quinebaugensis* (Figure 6).

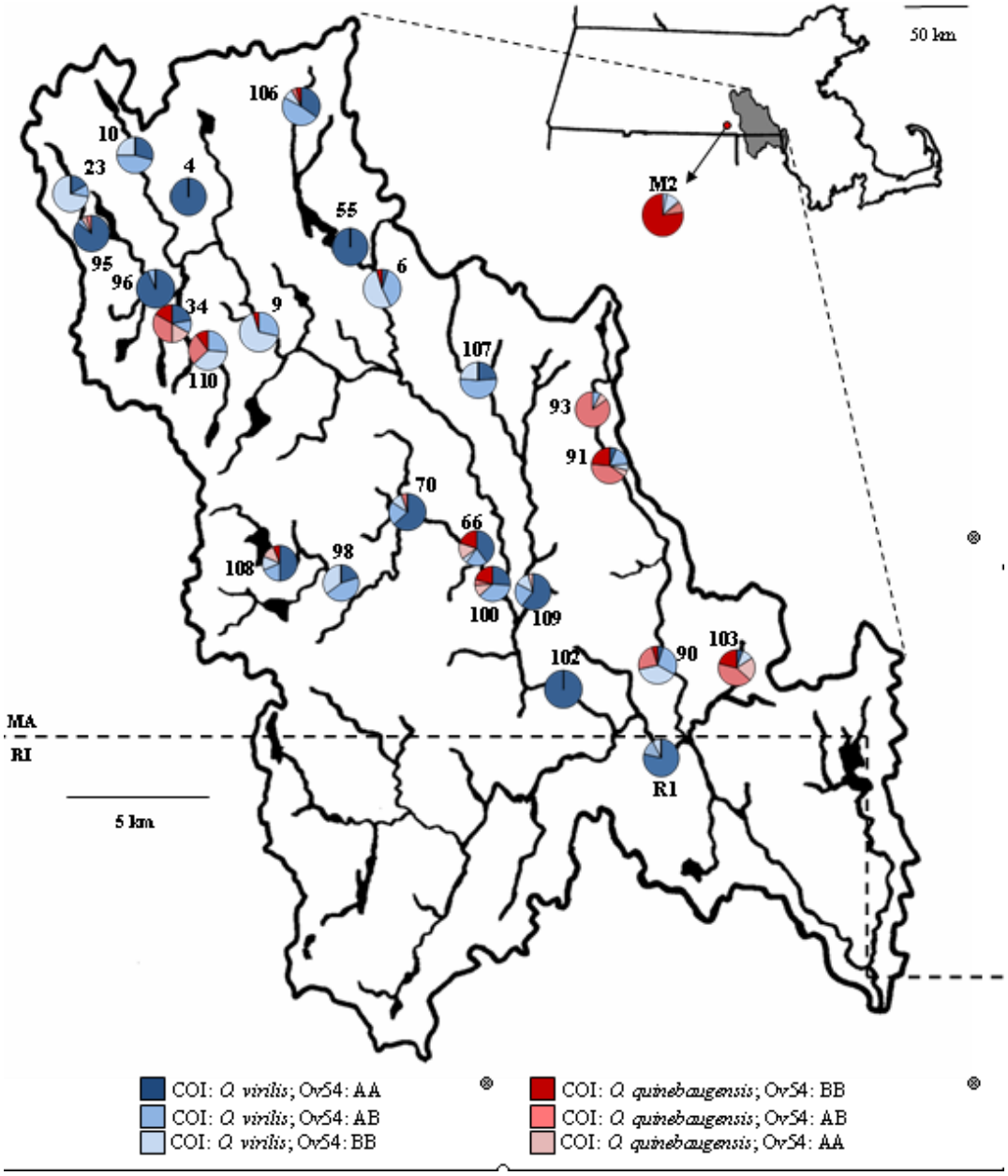

Figure 6: Collection Sites: This figure was taken from Mathews et al unpublished data. It shows the populations at various collection sites, including M2 and R1. Note how these two populations are unmixed compared to others with respect to COI haplotype and a nuclear microsatellite locus, Ov54. .

The inferences shown in Figure 6 were based on observations of mitochondrial DNA and form I male morphology of crayfish collected from these sites. Mitochondrial DNA was examined, and two distinct haplotypes (one for each species) were found in a 528-nucleotide stretch within the mitochondrial gene cytochrome oxidase I CO1. The difference within this region between

the M2 *O. quinebaugensis* group and the R1 *O. virilis* group is about 20 single nucleotide polymorphisms. The *O. virilis* group showed two haplotypes, but these differed by only a single SNP.

In order to assess how typical this COI haplotype was among *O. virilis* as a species, and not just among the *O. virilis* collected at the R1 site and other locations within the Blackstone River Valley, four *O. virilis* individuals from Michigan were examined. Their COI sites were also different from *O. quinebaugensis* by about 20 snps, differing with the other *O. virilis* samples by only a few nucleotides. This confirmed that the COI haplotypes were reliable at distinguishing the two species from each other (Mathews et al, unpublished data).

Behavioral differences between the two species were also observed. In the lab, when sexual partners of variable COI haplotypes were made available to collected crayfish on separate occasions, matings between members of different CO1 haplotypes were shown to be initiated less frequently than matings between similar COI haplotypes (Mathews et al, unpublished data).

At the R1 and M2 collection sites, where crayfish populations were assumed to be unmixed, strong statistical associations were observed between the CO1 haplotypes and two other traits: morphology and the genotype of a novel nuclear DNA microsatellite locus, Orco 54. In other collection sites within the Blackstone River Valley watershed, where the two species coexist, less strong associations between these traits were seen. Instead of traveling together, the genetic and morphological markers occur more independently. This seems to indicate transfer of genetic information from one species to another: hybridization may be occurring (Mathews et al, unpublished data).

Additional data from AFLP markers may shed more light on these populations. At present, full-genome analysis is not a feasible option for population studies like this one due to the high cost and labor required. Researchers must approximate genome-wide variation based on information from a limited number of loci (Bensch and Akkesson, 2005). Techniques that analyze microsatellites and SNPs have already been employed in the study of this problem.

These techniques can provide detailed genetic information about a few loci per experiment, but have the drawbacks of long start up and high cost of typing. Other drawbacks of these techniques are referred to above. AFLP overcomes some of these difficulties by being highly reproducible (unlike the high-throughput version of SNP analysis), and as a way to cheaply generate information about a plethora of loci at once. The time investment is therefore greatly reduced. As stated by Hardy (2003)"For population geneticists, the possibility of using RAPD or AFLP markers to assess relatedness between individuals and to study microgeographic isolation-by distance processes is promising because, compared to microsatellite markers, many polymorphic loci can be obtained fairly easily, in a relatively short time, and at a relatively low cost."

A drawback to AFLP analysis is that the per-locus amount of data is poor compared to microsatellite and SNP data. The latter two assays will give information regarding the locus's context within the genome, as well as the heterozygosity/ homozygosity of the marker. An AFLP analysis will give only a simple "present" or "not present" answer for a fragment. For a more in depth discussion of the AFLP method, see appendix.

Mariette et al (2002) provide some key insights when analyzing the problem of choosing between an assay that generates sparse data on a multitude of sites (AFLP) to assays that generate detailed information about a few sites (SNPs and microsatellites). According to this team, population studies on the same subjects often result in contradictory results when different molecular marker systems are used. In order to identify the possible cause of this inconsistency, they constructed computer simulations of various scenarios that could play out in an evolving population. The scenarios differed by population size, migration rate, and heterogeneity of gene flow. Once these virtual populations were constructed, researchers simulated the use of different marker systems to evaluate the genetic diversity within those populations. It was shown that a strong correlation existed between actual genetic diversity and the diversity estimate generated by the simulated assay, whichever marker system was used. This was particularly true in cases where genomic heterogeneity was low. However, it was found that in scenarios where this value was high, a reduced number of microsatellites

were insufficient to predict the diversity of the whole genome. In cases like this, more loci were needed to make an accurate prediction, and AFLP was found to be far more accurate.

The goal of this project was to uncover AFLP markers that can be used to track gene flow between the two species, if this is indeed occurring. Markers of this type would tend to be differentially present in the two species, for example, present in *O. virilis* and absent in *O. quinebaugensis.* Finding markers that display a differential pattern of this sort in areas where the two species do not intermingle, but that do not display the pattern in areas where the species co-exist would be evidence for hybridization, and would lend support to the hypothesis put forth by earlier researchers (Mathews personal communication, McMurough and Saltzman, 2009, Becker et al, 2009) that introgression is indeed occurring.

Materials and Methods

Samples & DNA Extraction

Genomic DNA from twelve different organisms was processed during this study. Four were from *O. Virilis* collected from R1, four were from *O. quinebaugensis* collected from M2, and four were from *O. virilis* collected from a site in Michigan. While the first two groups of four had already been extracted via methods similar to those described below, I extracted genomic DNA from the Michigan samples personally.

About 20 mg of tissue was acquired from gill slits and treated as laid out in the Solid Tissue Protocol in Gentra's Puregene: Genomic DNA Purification Kit, as summarized below.

First, each of the four tissue samples was submerged in 300µL of Cell Lysis Solution and 1.5µL of Puregene Protein Kinase K. together, these solutions work to rupture cell membranes and digest proteins, destroying the cell structure and dissolving the tissue. Small pestles were used to facilitate the liquefaction of the tissue. The suspended samples were vortexed thoroughly, and incubated over night at 55°C.

After about 12 hours, 150 uL of Protein Precipitation Solution was added to each sample. This solution causes proteins to crash out of solution, while leaving unbound DNA largely unaffected. Samples were vortexed, and then centrifuged at 13,000 gs for 10 minutes in a tabletop centrifuge. The protein precipitate was isolated in a yellow pellet at the bottom of the tube. Each supernatant was poured into separated tubes containing 300 uL of isopropanol. Suspension in a solution with high alcohol concentrations causes DNA to dissociate with water and crash out of solution. The samples were centrifuged gain at 13000 g until a white pellet was observed. The supernatants were poured off, and 300 uL of 70% ethanol was added to each sample. Samples were re-pelleted by centrifuging for 1 minute at 13,000 x gs. The alcohol supernatant was again discarded, and tubes were set on their sides to air dry at room temperature for about two hours. Finally, each DNA pellet was resuspended in 50µL of DNA Hydration Solution with vortexing. They were left to incubate overnight at 65°C. In the morning, samples were slowly cooled in a Styrofoam box, then frozen at -20°C. Later, they would be melted at 37°C before each use.

Using known concentrations of lambda DNA for comparison, small aliquots of each extraction were mixed with Syber Green and glycerol and were run out on a 2% agarose gel. Their concentrations were found to be about 200 ng/uL.

AFLP analysis: Digestion-ligation reactions

The AFLP protocol used in this experiment is based on a protocol posted by Professor Paul Wolf, from the Department of Biology department of Utah State University. This protocol is available at http://bioweb.usu.edu/wolf/aflp_protocol.htm and is cited below.

In essence, each of the twelve samples was amplified using three different sets of AFLP primers, giving a total of thirty-six banding patterns.

As described in the appendix, adaptor sequences are necessary for this procedure. The adaptors are prepared in the following manner: complementary nucleotide sequences were prefabricated by IDT integrated DNA technologies. The adaptors pairs (two complementary strands relevant to the same cut site) were mixed so to obtain a final concentration of 5uM for EcoRI adapter pair and 50uM for the MseI adaptor pair. Because MseI is a four-base cutter and

EcoRI is a six-base cutter, MseI will tend to leave 16 times as many sticky ends for adaptors to bind to than will EcoRI, so using a higher concentration of the MseI adaptor is appropriate. Then, the adaptors were heated at 95 ˚C for 5 min to denature. Then, they were allowed to cool slowly in a Styrofoam box to renature completely. They were frozen at -20˚C, and thawed at 37˚C before each later use.

Next, aliquots of genomic DNA from each of the twelve organisms were digested using MseI and EcoRI. An enzyme master mix of the following proportions was prepared:

- NaCl 0.5M 0.1µl
- BSA @ 1mg/ml 0.05µl
- MseI: 1 unit 0.1µl
- EcoRI: 5 units 0.12µl
- Deionized water 0.33µl

Each of the twelve samples was digested by adding the following to each tube:

- 2.5 µl NEB 2 buffer
- 1µl EcoR1/MseI master mix
- 50ng template DNA (about .25 uL)
- Distilled water to a total volume of 12.5µl

These reactions were left to incubate at 37˚C for a minimum of two hours. Next, the pre-mixed adaptor sequences were ligated to the DNA fragments from the previous reaction by adding the following to each of the twelve previous reaction tubes:

- 12µl adaptor/ligation solution

- 0.5µl T4 DNA Quick Ligase
- 0.5 uL Quick Ligase Buffer
- For a total now of 25.5µl per tube.

These were incubated at room temperature for about two hours. Afterwards, they were frozen at -20˚C, and later melted at 37˚C.

AFLP analysis: Selective PCR cycles

Once the adaptor-ligated fragments were prepared, two rounds of PCR were used to selectively amplify certain fragments from the mixture. The first PCR reaction uses primers that match the adapter sequence and have one additional selective base. Theoretically, this could be any of the four bases. I used two sets of selective primers for each of the twelve samples: Mse-C with Eco-A, and Mse-A with Eco-C. This produced a total of twenty-four reactions.

Tubes were prepared with the following contents:

- 1X Taq Buffer
- .25U DNA polymerase
- 1.5 mM MgCl2
- 0.12mM dNTP mix
- 0.2µM Primer-Mse-C
- 0.2µM Primer-Eco-A
- 2.5 µL of the tenfold diluted product of the previous reaction (for which no accurate concentration can be projected)

The final volume of reaction was be 25 µL. Tubes were placed in a thermo cycler using the following parameters:

1. 72 ˚C 2min
2. 94 ˚C 30 sec
3. 56 ˚C 30 sec
4. 72 ˚C 2 min
5. Goto 2 29 more times
6. 60 ˚C 10 min
7. 25 or 4 ˚C hold

Tubes were held at 4˚C for as many as 24 hours.

The second round of selective PCR was set up similarly to the first, with some minor alterations. First, instead of using the digestion-ligation reaction as a template as above, a ten-fold diluted sample of the previous PCR product was used. This sample had to be diluted to prevent excess binding of the primers from the previous cycle to the template, which would have resulted in dNTPs being used to make unreadable fragments.

The second change was to add a different set of primers. The second PCR reaction uses primers that match the adapter sequence and have three additional selective bases. Theoretically, the first of these bases must match the selective base used in the first PCR, while the second two bases could be any base. I used three sets of selective primers for each of the twelve samples. The EcoRI primers in this second selection were all labeled with 6-FAM, to make them visible for fragment analysis later. Non-labeled fragments would be invisible. The twelve samples that had been amplified with Mse-A and Eco-C, were amplified a second time with Mse-ATC and Eco-CTC in one tube, and Mse-ATC and Eco-CAG in another tube resulting in twenty-four finished samples. The twelve samples that had been amplified with Mse-A and Eco-A were

amplified a second time with Mse-ATC and Eco-ACG, resulting in twelve finished samples. Together, this made thirty-six distinct, finished samples. Once these thirty-six samples were made, 1 uL from each was suspended in separate PCR tubes with 9 uL of Hi-Dye formamide, and 0.5 uL Genescan-500LIZ size standard. The tubes were labeled and sent to the Sequence Analysis Facility on Science Hill at Yale University. The .fsa files that were produced from this analysis were then analyzed using Genemapper 4.0 (Applied Biosystems, Inc).

The markers were located by reading the .fsa files generated by the Yale fragment analysis facility using the program GeneMapper version 4. These .fsa files contained peak information in the form of line graphs for each of the thirty-six samples. The peak information was generated in the following manner: as the labeled fragments in the electrophoresis gel, which was produced by running out the samples as described in the materials and methods section above, were run past a photo-recording device at the Yale fragment analysis facility, the florescent 6-FAM primers used in the second PCR cycle of AFLP caused an increase in the detection of relative fluorescence units (RFUs, the y axis of the line graphs), resulting in a peak. The absence of a fragment of a particular length in base pairs would appear as the absence of a peak at that point in the gel. While the original x axis of the .fsa files reflected the physical location of a fragment on the original gel ass a timestamp, the moment when the fragment passed the photorecording device. GeneMapper was used to convert this time value into length in base pairs, relative to the GS LIZ 600 size standard that was multiplexed with each sample. The result was 36 line graphs with RFUs as the y-axis, length in base pairs as the x-axis, and a single line representing the detected amount of fluorescence for each sample amplified by one set of primers. 6 graphs that failed to normalize, even after manual adjustment of the size match editor, were labeled contaminated and disregarded from further analysis. To avoid mislabeling non-peaks as peaks, the analysis method was set by the user to disregarded peaks below 200 RFUs. However, if these peaks were judged by eye to be part of an obvious (see screen shots below) marker, they were recorded manually.

Variable markers were detected by overlaying peak data from DNA samples taken from different organisms that were amplified using the same primer sets. In the overlay panel, each

of the collection sites was assigned a color, to allow for easy manual recognition of discrepancies between species groups. Potential markers were flagged in areas where a peak was recorded for all of the M2 *O. quinebaugensis,* and not recorded for all of the R2 and MI *O. virilis* (or vice versa). In these cases, a screen capture was taken, and the allele's size, designation (automatically assigned by GeneMapper), the organisms involved, and the primer set used to produce the fragment were recorded. The software was also used to calculate the peak heights and the area beneath individual peaks.

Results and Discussion

This section catalogs and discusses the results of the search for AFLP markers that correlate with species identity between *O. virilis* and *O. quinebaugensis*. Twelve marker candidates were found, as summarized in the Table 1. Each marker is defined by a fragment which is consistently present or absent in all of theorganisms from one species, and not in the other. So, for example, a fragment that is present in all four O. quinebaugus samples, and also in one of the eight O. virilis samples would not be considered a marker by the selection standards applied to this data. In other words, no variations within the populations were tolerated.

Table 1. Total Markers Detected-Populations in which markers were found to be present or absent without variation. This table lists all twelve of the fragments that were found to be differentially present or absent between the two species. The figures and tables below and in the appendix all elaborate on data that is summarized here.

Allele designation	Length (bp)	Primer Set	M2?	Mich?	R1?
15	82	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
16	83	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
21	94	Mse1-ATC: Eco R1-ACG	Yes	No	No
22	95	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
53	156	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
75	191	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
123	419	Mse1-ATC: Eco R1-ACG	No	Yes	Yes
45	154	Mse1-ATC: Eco R1-CAG:	No	Yes	Yes
50	162	Mse1-ATC: Eco R1-CAG:	Yes	No	No
110	271	Mse1-ATC: Eco R1-CAG:	No	Yes	Yes
145	343	Mse1-ATC: Eco R1-CAG:	Yes	No	No
171	509	Mse1-ATC: Eco R1-CAG:	No	Yes	Yes
57	171	Mse1-ATC: Eco R1-CTC	No	Yes	Yes
119	362	Mse1-ATC: Eco R1-CTC	No	Yes	Yes

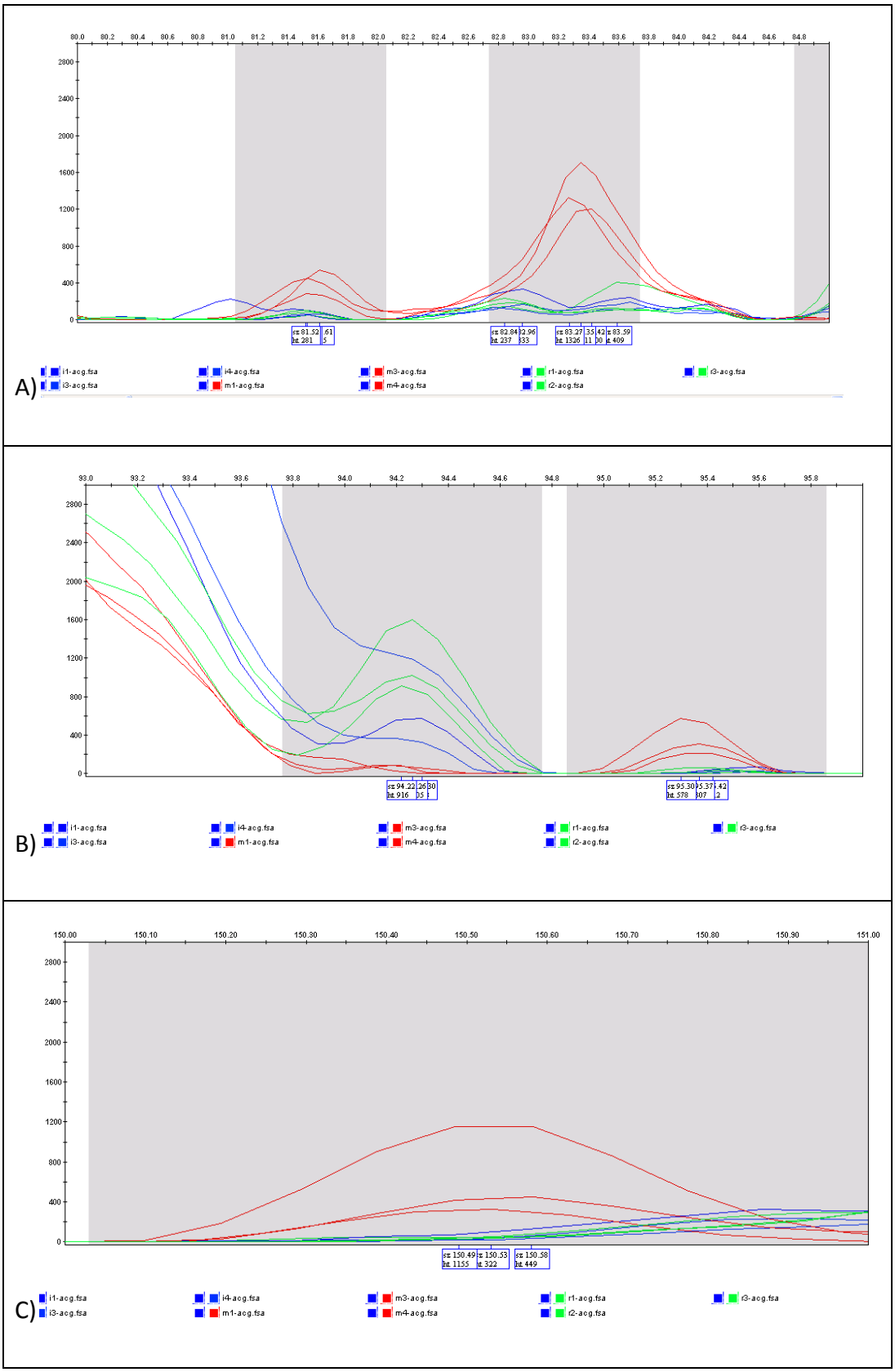
A)
i1-acg.fsa
i3-acg.fsa
i4-acg.fsa
m1-acg.fsa
m3-acg.fsa
m4-acg.fsa
r1-acg.fsa
r2-acg.fsa
r3-acg.fsa
B)
i1-acg.fsa
i3-acg.fsa
i4-acg.fsa
m1-acg.fsa
m3-acg.fsa
m4-acg.fsa
r1-acg.fsa
r2-acg.fsa
r3-acg.fsa
C)
i1-acg.fsa
i3-acg.fsa
i4-acg.fsa
m1-acg.fsa
m3-acg.fsa
m4-acg.fsa
r1-acg.fsa
r2-acg.fsa
r3-acg.fsa

Figure 7. Five of the twelve markers shown as chromatographs. In these figures, peaks correspond to the brightness given off by the presence of labeled primers, which translates to the presence of a fragment. The X axis describes the size of the fragment in base pairs, while the Y axis describes the strength of the signal in relative fluorescence units, which translates to the number of copies of the fragment. The area under the peak is also an expression of the strength of the signal. Panel A is Alleles 15 and 16 for Primer pair MseI 1-ATC: Eco R1-ACG. In this panel, the grey bin to the left demarks allele 15 at around 82 bp, while the grey bin to the right demarks allele 16 at around 83 bp. In both cases, organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-ACG. Panel B shows Alleles 21 and 22 for Primer pair Mse1-ATC: Eco R1-ACG. In this panel, the grey bin to the left demarks Allele 21 at around 94 bp, while the grey bin to the right demarks Allele 22 at around 95 bp. In the case of allele 21, organisms from the M2 collection site (marked in red) are shown not to contain the fragment, while organisms from MI and R1 (blue and green) contain the fragment. In the case of allele 22, organisms from the M2 collection are shown to contain the fragment, while organisms from MI and R1 do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-ACG.Finally, in panel C, one can see Allele 53 for Primer Pair Mse1-ATC: Eco R1-ACG In this panel, the grey bin demarks Allele 53 at around 156 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-ACG. The tables below provide more detailed information about these peaks.

Table 2. Alleles 15, 16, 21, 22, and 53 for Primer pair MseI 1-ATC: Eco R1-ACG. This table elaborates on the preceding figure by listing precise fragment sizes, peak heights, and peak areas. The heights and peaks areas both describe the strength of the signal, which translates to the number of copies of the fragment. As stated previously, heights below 200 RFUs were disregarded as being too small to count as a definitive positive signal. Chromatographs showing peaks of less than 200 RFUs show the ambiguity of those peaks verses the baseline: it is not clear whether a signal is being emitted. "Organism" in the table below indicates which organism the fragment was detected in. M2, Mich, and R1 are the collection sites from which the organisms were harvested, and the additional numeral, 1, 2, 3, or 4 indicates which one of four organisms the peak was detected in. One may notice that some samples are excluded. This is because the chromatographs of thos samples failed to normalize, which may have been due to contamination of the original sample.

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
15	Mich-1	na	<200	na
15	Mich-3	na	<200	na
15	Mich-4	na	<200	na
15	M2-1	81.61	545	28580
15	M2-3	81.53	449	2779
15	M2-4	81.52	281	1420
15	R1-1	na	<200	na
15	R1-2	na	<200	na
15	R1-3	na	<200	na
16	Mich-1	na	333	na
16	Mich-3	na	<200	na
16	Mich-4	na	<200	na
16	M2-1	83.42	1200	9946
16	M2-3	83.27	1327	11237
16	M2-4	83.35	1711	13122
16	R1-1	na	<200	na
16	R1-2	na	409	na
16	R1-3	na	237	na
21	Mich-1	94.3	578	2838

21	Mich-3	94.2	420	na
21	Mich-4	94.25	400	na
21	M2-1	na	<200	na
21	M2-3	na	<200	na
21	M2-4	na	<200	na
21	R1-1	94.26	1025	5860
21	R1-2	94.26	1605	9101
21	R1-3	94.22	916	4580
22	Mich-1	na	<200	na
22	Mich-3	na	<200	na
22	Mich-4	na	<200	na
22	M2-1	95.37	307	1206
22	M2-3	95.3	578	2259
22	M2-4	95.42	212	800
22	R1-1	na	<200	na
22	R1-2	na	<200	na
22	R1-3	Na	<200	na
53	Mich-1	Na	<200	na
53	Mich-3	Na	<200	na
53	Mich-4	Na	<200	na
53	M2-1	150.53	322	1452
53	M2-3	150.58	449	2365
53	M2-4	150.49	1155	5741
53	R1-1	Na	<200	na
53	R1-2	Na	<200	na
53	R1-3	Na	<200	na

As can be seen in the figure 8 and Appendix B, some very sharp discrepancies have been detected between the peak data from individuals collected from sites R1, Mich, and M2. However, these differences require further validation with increased specimens in order to determine if they can be considered useful AFLP markers. Testing a larger pool of organisms for the presence and absence of these twelve fragments could either confirm or deny their association with species identity. Further testing of this type is recommended.

The marker candidates listed in Table 2 can be used as a test of the three primer sets that were used to generate this data, meaning that it is likely that not all twelve of the marker candidates will prove to have consistent associations with species identity in non-hybrids. A marker candidate would be disqualified by using the same primer set to perform AFLP on genomic DNA from a new organism, either an *O. virilis* or *O. quinebaugensis*, and finding that the peak data generated in this analysis is not consistent with associations highlighted in the original twelve organisms. Even one exception in one non-hybrid organism would be enough to disqualify a marker as one with 100% accuracy. However, compromises may have to be made, since the M2 and R1 populations were only nearly pure, meaning that some hybrids were present. One can say with certainty that no more than twelve markers with 100% accuracy will be found by amplifying *O. virilis* and *O. quinebaugensis* DNA using these particular primer sets. This is because exceptions to the association of the presence or absence of a fragment with species identity of any additional markers have already been found in this peak data, and can be observed in the original .fsa files. These exceptions are why those fragments were not flagged as candidates in the original analysis.

Once these twelve markers are tested among a larger pool of organisms, it may be established that some of them are indeed AFLP markers that can accurately distinguish between non-hybrid members of the different species. Between 30 and 50 genomes are available from each group for testing. Once this is established, the mixed occurrence of markers that would normally occur separately would be evidence of a hybridization event. For example, an organism which contained both of the fragments highlighted in Figure9, designated alleles 21 and 22, would contain genetic information from both species.

The location and abundance of such hybrid organisms could imply various hypothesis about gene flow in the BRV watershed. For example, a complete absence of hybrids might either negate the hypothesis that introgression is occurring, or suggest that the traits which the markers are associated with are somehow deleterious when transposed from one species to another. An overabundance of such hybrids might bolster the hybridization hypothesis, and imply that it is well underway. Hopefully, at least some of the twelve markers that were discovered during the course of this research will be useful for determining the activity of these two populations.

Works cited

Avise, J C. Phylogeography: The History and Formation of Species. Cambridge: Harvard University Press, 2000.

Avise, J C, C A Reeb and N C Saunders. "Geographic population structure and species differences in mitochondrial DNA of mouthbrooding marine catfishes (Ariidae) and demersal spawning toadfishes (Batarachoididae)." Evolution 41 (1987): 991-1002.

Ayres, D R, et al. "Extent and degree of hybridization between native and exotic Spartina cordgrass in California, USA, determined by RAPD." Molecular Ecology 8 (1999): 1179-1186.

Bartlett, J and M Stirling. PCR Protocols. 2. Totowa: Human Press, 2003.

Becker, M, T Mulhern, and M Perry. "Introgressive Hybridization between Native and Invasive Crayfish: A Test of Reproductive Isolation." WPI Major Qualifying Project.

Bensch, S and M Akesson. "Ten years of AFLP in ecology and evolution: Why so few animals?" Molecular Ecology 14 (2005): 2899-2914.

Cooper, R M and J P Zubeck. "Effects of enriched and restricted early environments on the learning ability of bright and dull rats." Canadian Journal of Psychology 12 (1958): 159-164.

Diekmann, K, et al. "Complete chloroplast genome sequence of a major allogamous forage species, perrenial ryegrass (Lolium perenne L.)." DNA Research 16.3 (2009): 165-176.

Dong, S and E Wang. "Flexible use of high-density oligonucleotide arrays for single-nucleotide polymorphism discovery and validation." Genome Research 11 (2001): 1418-1424.

Emery, N and N Clayton. "The mentality of crows: Convergent evolution of intelligence in corvids and apes." Science 306.5703 (2004): 1903-1907.

Freeland, J. Molecular Ecology. 1. West Sussex: John Wiley and Sons, 2006.

"Orconectes." Hamr, P. Biology of Freshwater Crayfish. Ed. D M Holdich. Oxford: Blackwell Science Ltd, 2002.

Harris, H. "Enzyme polymorphisms in man." Proceedings of the Royal Society of London Series B 164 (1966): 298-310.

Holeck, K T, et al. "Bridging troubled waters: Biological invasions, transoceanic shipping, and the Laurentian Great Lakes." BioScience 54 (2004): 919-929.

Jara, L, et al. "Distribution of (CGG)n and FMR-1 associated microsatellite alleles in a normal Chilean population." American Journal of Medical Genetics 75.3 (1998): 277-282.

Jeffreys, A J, V Wilson and S L Thein. "DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees." American Journal of Human Genetics 39.1 (1994): 11-24.

Jenkins, M, R Green and J Madden. "The challenge of measuring global change in wild nature: Are things getting better or worse?" Conservation Biology 17.1 (2003): 20-23.

Kiem, P, M Ert and T Pearson. "Anthrax molecular epidemiology and forensics: Using the appropriate marker for different evolutionary scales." Infection, Genetics, and Evolution 4.3 (2004): 205-213.

Kolar, C S and D Lodge. "Freshwater nonindigenous specis: Interactions with other global changes." Mooney, H A and R J Hobbs. Invasive Species in a Changing World. Washington: Island Press, 2000. 3-30.

Lewontin, R C and J Hubby. "A molecular approach to the study of genetic heterozygosity in natural populations. ||. Amounts of variation and degree of geterozygosity in natural populations of Drosophilia psuedoobscura." Genetics 54 (1966): 595-609.

Lynch, M and B G Milligan. "Analysis of population genetic structure with RAPD markers." Molecular Ecology 3.2 (1994): 91-97.

Mariette, S, et al. "Sampling within the genome for measuring within-population diversity: Trade-offs between markers." Molecular Ecology 11 (2002): 1145-1156.

Mathews, L and A Warren. "A new crayfish of the genus Orconectes Cope, 1872 from southern New England (Crustacea: Decapoda: Cambariidae)." Proceedings of the Biological Society of Washington 121.3 (2008): 374-381.

Mayr, E. Systematics and the Origin of Species. New York: Columbia University Press, 1942.

McMurrough, M R, Saltzman, L. "Introgressive hybridization between native and invasive crayfish: a study in the blackstone river valley." WPI Major Qualifying Project.

Mullican, M E. The Replication Process. 1996. 10 January 2009 <http://www.mun.ca/biochem/courses/3107/Topics/Replication_process.html>.

Neal, D. Introduction to Population Genetics. Cambridge University Press, 2004.

Oliveira, E J, et al. "Origin, evolution and genome distribution of microsatellites." Genetics and Molecular Biology 29 (2006): 294-307.

Padilla, D K and S L Williams. "Beyond ballast water: Aquarium and ornamental trades as sources of invasive species in aquatic ecosystems." Frontiers in Ecology and the Environment 2 (2004): 131-138.

Polymerase Chaine Reaction. 1 February 1998. <http://www.gene.com/ae/AB/GG/polymerase.html>.

Ricciardi, A and H J MacIsaac. "Recent mass invasion of the North American Great Lakes by Ponto-Caspian species." Trends in Ecology and Evolution 15 (2000): 62-65.

Ryhmer, J and R Simberloff. "Extinction by hybridization." Annual Review of Ecology and Systematics 27 (1996): 83-109.

Schaal, B, K Olsen and J CB Carvalho. "Evolution, domestication, and agrobiodiversity in the tropical crop cassava." Motley, T J, N Zerega and H B Cross. Darwin's harvest: New approaches to the origins, evolution, and conservation of crops. New York: Columbia University Press, 2006. 269-284.

Sia, E A, et al. "Analysis of microsatellite mutations in the mitochondrial DNA of Saccharomyces cerevisiae." Proceedings of the National Academy of Sciences 97 (2000): 250-255.

Swallow, J G, et al. "Selection experiments and experimental evolution of performance and physiology." Experimental Evolution 12 (2009): 301-351.

Trevor, G and B Rowe. An Introduction to Molecular Ecology. 2. Oxford: Oxford University Press, 2008.

Wolf, D, N Takebayashi and L Rieseberg. "Predicting the risk of extinction through hybridization." Conservation Biology 15 (2001): 1039-1053.

Wolf, J. "AFLP Protocol." *University of Utah* 4 Sept 2000: n. pag. Web. 29 Apr 2010. <http://bioweb.usu.edu/wolf/aflp_protocol.htm>.

Appendix:

A- AFLP: Basic Principles

The initiation of the AFLP protocol is similar to a simple digestion reaction: the genomic DNA in question is digested with a pair of restriction enzymes. The two enzymes conventionally used are *EcoR1* and *Mse1*. These enzymes are a six-cutter and a four-cutter respectively, meaning that they restrict DNA at sequence specific sites either six of four nucleotides long. *EcoR1* cuts at (5'-G*AATTC-3'), while *Mse1* cuts at (5'-T*TAA-3') (Bensch and Akesson, 2005).

The digestion of genomic DNA with these two restriction enzymes generates hundreds of thousands of DNA fragments of various lengths. Each of the fragments will have sticky ends, overhanging single-stranded DNA at the tips of the fragment. Each sticky end is unique to the restriction enzyme that left it. For example, the sticky end left by *Mse1* would be as follows:

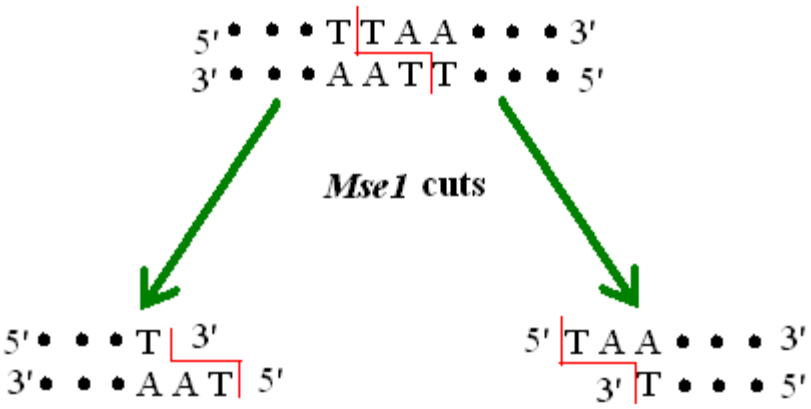

Figure8. Sticky End Formation

In this figure, *Mse1* leaves a pair of identical sticky ends after cutting DNA at a recognized restriction site.

The uniqueness of the sticky ends results in three possible types of fragments: *Mse1-Mse1*, *Mse1-EcoR1*, and *EcoR1-EcoR1* (Bensch and Akesson, 2005).

In the second stage of the AFLP protocol, two types of adaptor sequences are ligated to the tips of the DNA fragments. These adaptor sequences are short, double-stranded, and terminate with sticky ends that are complementary to one of the two possible sticky ends left by the restriction enzymes. The ligation procedure joins the adaptors with DNA fragments in a chemical reaction that creates a continuous DNA strand. In this new strand, the restriction sites are not reformed, because the adaptor sequences contain nucleotides immediately adjacent to the sticky ends that are not part of the restriction enzyme's recognized sequence (Bensch and Akesson, 2005). For example, the sticky end of a DNA fragment left by *Mse1* might be ligated to an adaptor sequence in the following way:

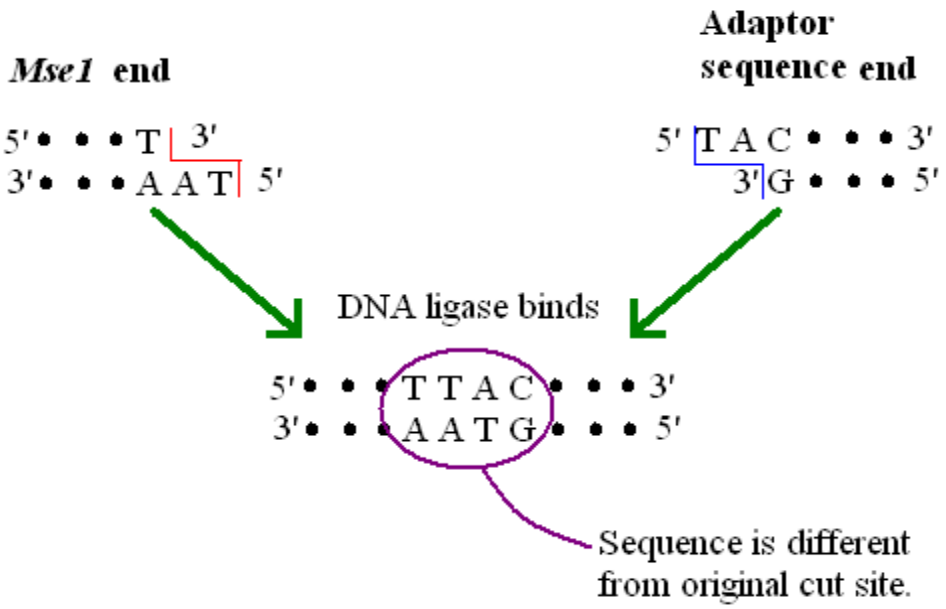

Figure 9. Cut Site Elimination **In this figure, the ligation of a hypothetical adaptor sequence to a sticky end left by *Mse1* results in a sequence that cannot be recognized and cut by *Mse1*. Because of this cut site elimination, the restriction and ligation reaction steps can be run in the same tube.**

The chemistry of a ligation reaction is shown in Figure. 10.

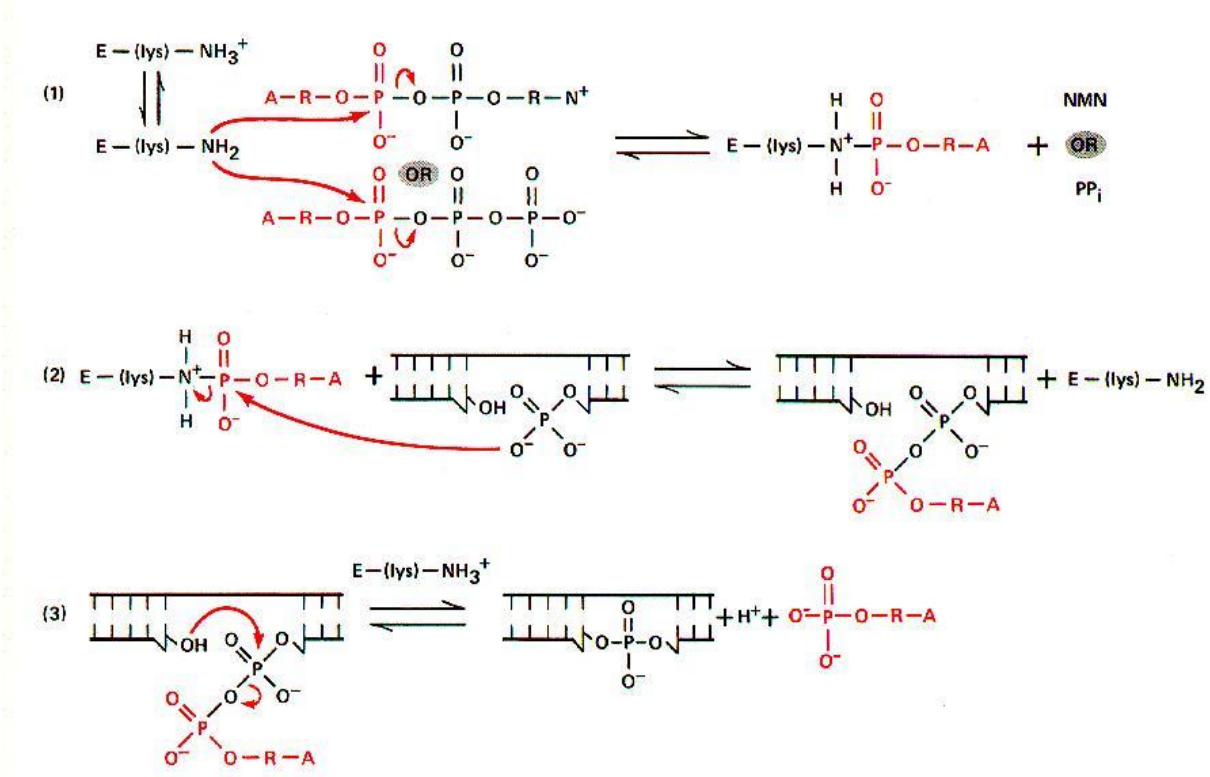

Figure10: Ligation Reaction This figure, taken from Martin E. Mulligan's "Replication Process" at http://www.mun.ca/biochem/courses/3107/Topics/Replication_process.html, shows how Ligase joins DNA fragments. (1)First, an enzyme-nucleotide intermediate is formed by transfer of the adenylyl group from ATP or NAD (cofactors) to the epsilon amino group of a lysine residue in the active site of the enzyme, DNA Ligase. (2)Second, the adenylyl group is transferred from the enzyme to the 5'phosphate group in the nick in the phosphate backbone of the DNA to be ligated. This creates a pyrophosphate linkage and a good leaving group. (3) Finally, nucleophilic attack by the 3' hydroxyl group on the phosphorus atom causes the displacement of adenylyl group and formation of a new phosphodiester bond in the DNA, now a continuous chain. In the case of joining sticky ends, this reaction would happen twice, once on each strand.

The third and fourth phases of the AFLP protocol require the use of the Polymerase Chain Reaction, or PCR. PCR is a method of creating many copies of a DNA molecule using the original molecule as a template. This process is divided into three main steps. First, the double-stranded DNA of interest is melted at about 95°C into single strands. Second, the temperature of the reaction is lowered to about 65°C so that a pair of primers can anneal to the single strands. Primers are special sequences that anneal to short, complementary sequences at the ends of the region of interest within the template DNA. The temperature is then raised to about 72°C, the optimal operating temperature for the enzyme, Taq polymerase. Taq is one of many DNA

polymerases, but it is typically used because of its ability to retain viability at DNA melting temperatures. At this point in the reaction, the annealed primers direct the Taq polymerase to sequence the complementary DNA strands from free nucleotides. Taq polymerase travels along a single stranded DNA molecule, adding complementary bases as it travels, which results in a double stranded DNA molecule. The cycle is then repeated: the new DNA molecule is melted at 95', primers anneal at 65°C, and Taq elongates the new strand at 72°C. This procedure is repeated many times, creating exponentially more DNA strands with each cycle (Bartlett, 2003). The exponential nature of the PCR reaction is exemplified in Figure below:

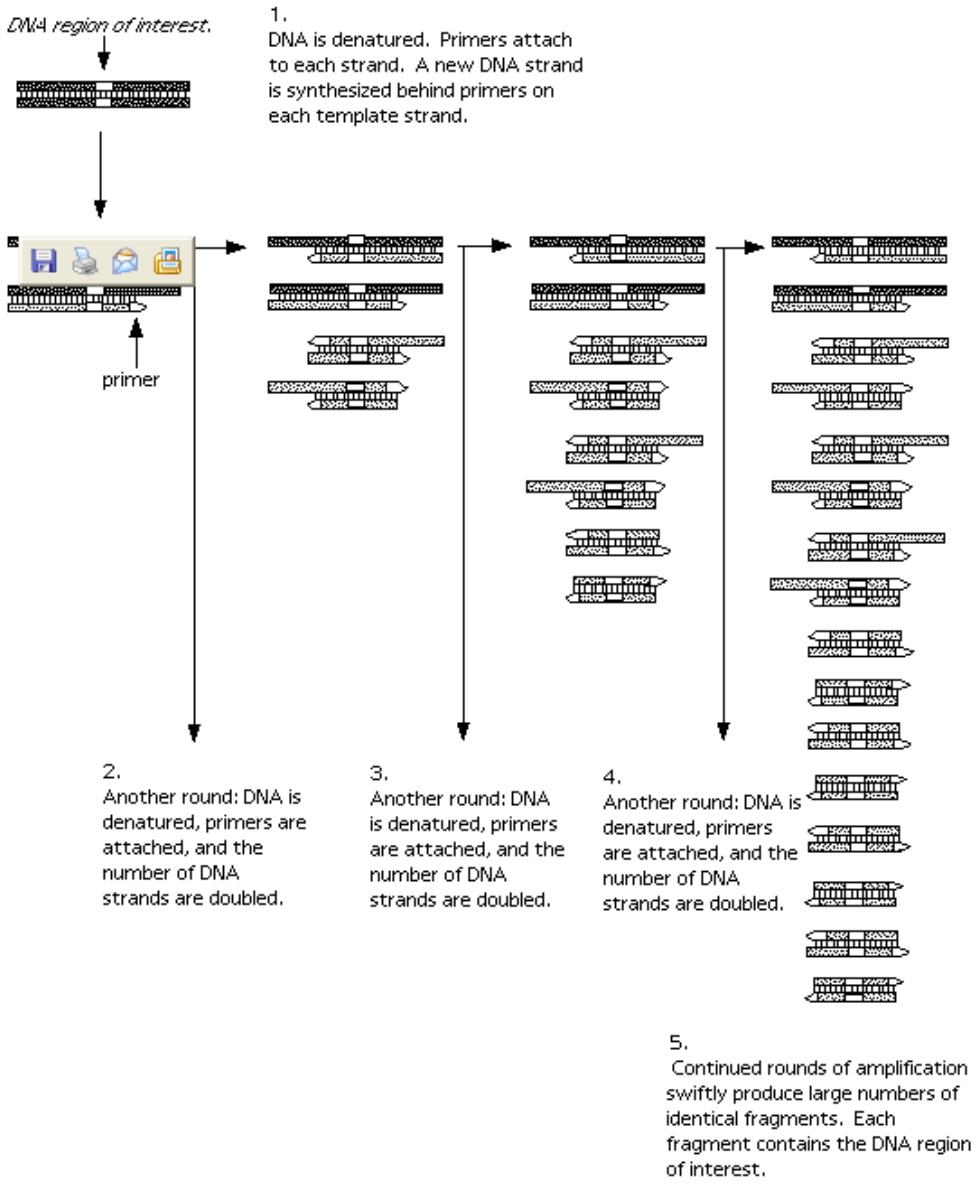

Figure 11: Target DNA Copied Faster than Non-Target DNA in PCR

This figure from Biotech Graphic Gallery, 1998, available at http://www.gene.com/ae/AB/GG/polymerase.html shows the exponential nature of the PCR amplification process. Each cycle has more templates from which to copy, so that each cycle creates more copies than the last. This diagram also shows the exclusivity of the reaction provided by the primers. In cases where the target sequence is buried in a region of other DNA, primers adjacent to the region of interest are applied, ensuing that this region will be copied with greater frequency than non-target regions.

The third and fourth phases of the AFLP protocol use PCR in a way reminiscent of RAPD. Arbitrarily selected primers reduce the myriad of fragments in two steps. First, in a preamplification step, primers are applied that are complementary to the ligated adaptor sequences plus one additional nucleotide, which could be A, C, G, or T. Only the fragments exhibiting this sequence information will be amplified in the PCR reaction. The fragment population which results from the PCR reaction has only 1/16 of the diversity of the original population (¼* ¼). A second, even more selective amplification step uses a small aliquot of this population as a template. The primers used in this step are complementary to the adaptor sequence, plus the additional nucleotide from the previous step, plus two more arbitrarily selected nucleotides. This further reduces the number of fragments by 1/256 (¼* ¼ *¼* ¼). One of the two primers in this step, typically the one complementary to the fragment end that contains the remnant of the *EcoR1* sticky end, is fluorescently labeled (Bensch and Akesson, 2005).

In the final step of the AFLP protocol, the fragments produced by selective amplification are separated from each other via DNA gel electrophoresis. This process entails forcing DNA fragments, which carry a negative charge due to their multiple phosphate groups, through a matrix of agarose gel by way of an applied electrical field. The matrix encumbers the passage of DNA, more so for larger fragments, so that smaller fragments run exponentially farther in comparison to longer fragments. The travel of linear fragments is always directly proportional to the log of the fragments' length, so that the size of fragments can be estimated by measuring the distance they travel. Usually, a ladder sample containing multiple fragments of known sizes is run alongside the experimental sample, to serve as a comparison (Bensch and Akesson, 2005)

Three types of fragments have been amplified,*Mse1-Mse1, Mse1-Ecor1,* and *EcoR1-EcoR1.* Because only one type of primer in the second amplification step contains a fluorescent label, only 2/3 of the bands on the final gel will be visualized. These will be the bands that contain fragments with the fluorescent label. This selection further reduces the population of DNA for analysis (Bensch and Akesson, 2005).

B- Additional Data

This section contains data about the seven markers not discussed in depth in the discussion section above. For notes on how to interpret the data, see similar headings above.

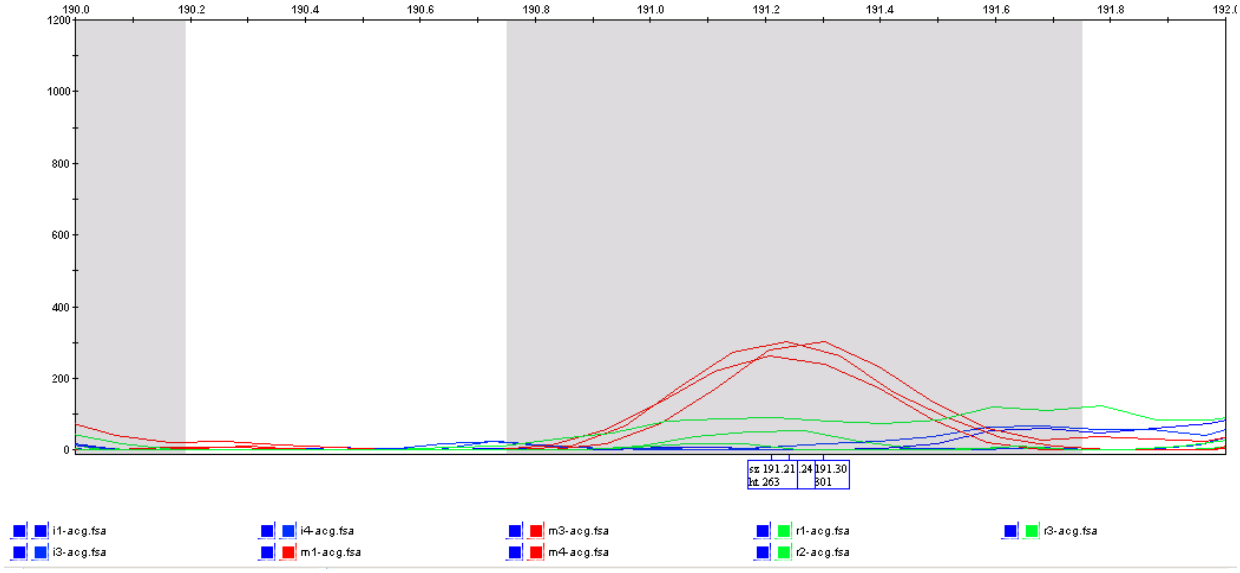

Figure12. Allele 75 for Primer Pair Mse1-ATC: Eco R1-ACG

In this figure, the grey bin to the right demarks Allele 75 at around 191 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-ACG.

Table 3. Allele 75 for Primer Pair Mse1-ATC: Eco R1-ACG This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
75	Mich-1	Na	<200	na
75	Mich-3	Na	<200	na
75	Mich-4	Na	<200	na
75	M2-1	191.24	303	1391
75	M2-3	191.3	301	1403
75	M2-4	191.21	263	1212
75	R1-1	Na	<200	na
75	R1-2	Na	<200	na
75	R1-3	Na	<200	na

.

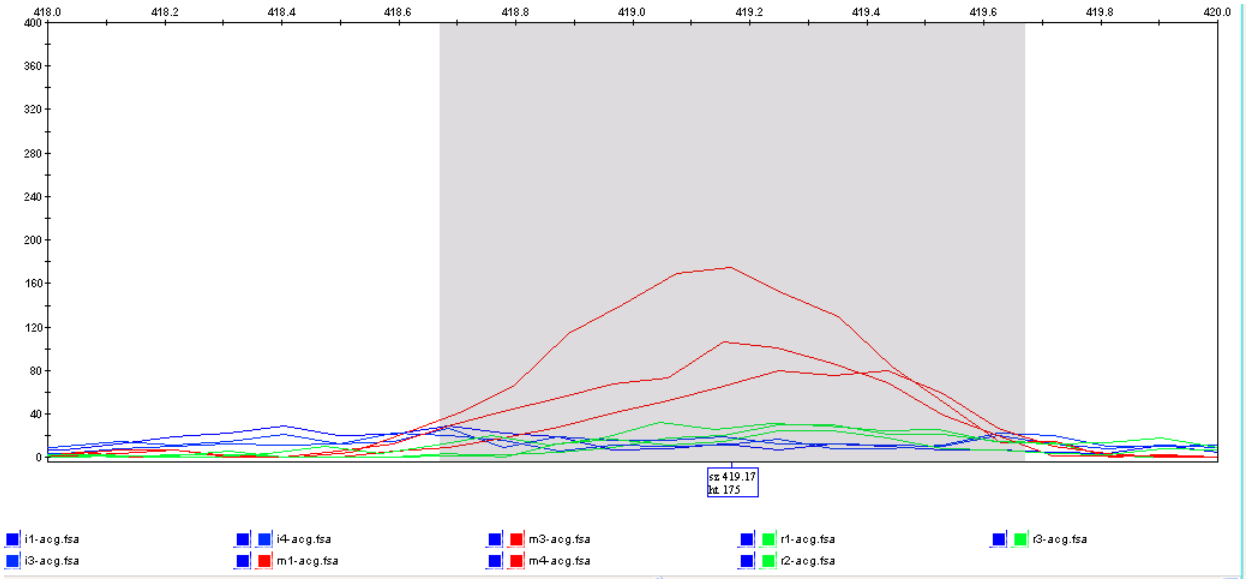

Figure 13. Allele 123 for Primer Pair mse1-ATC: Eco R1-ACG

In this figure, the grey bin demarks Allele 123 at around 419 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-ACG.

Table 4. Allele 123 for Primer Pair mse1-ATC: Eco R1-ACG

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
123	Mich-1	Na	na	na
123	Mich-3	Na	na	na
123	Mich-4	Na	na	na
123	M2-1	419.17	201	1187
123	M2-3	419.17	120	na
123	M2-4	419.2	100	na
123	R1-1	Na	na	na
123	R1-2	Na	na	na
123	R1-3	Na	na	na

This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.

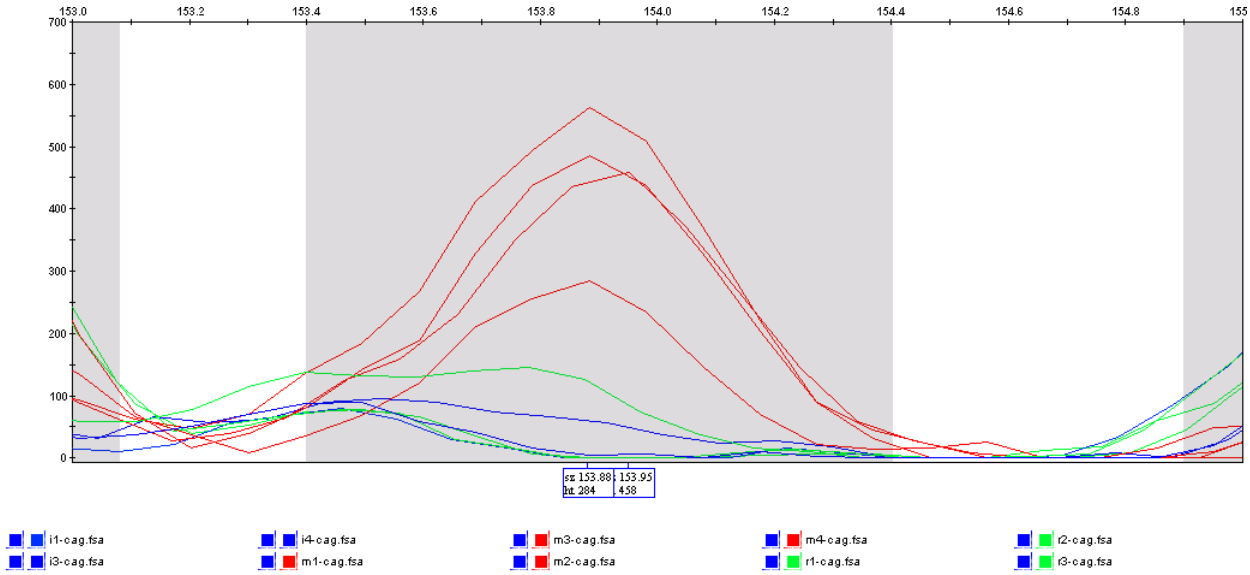

Figure 14. Allele 45 for Primer Pair Mse1-ATC: Eco R1-CAG:

In this figure, the grey bin in the center demarks Allele 45 at around 154 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CAG.

Table 5. Allele 45 for Primer Pair Mse1-ATC: Eco R1-CAG:

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
45	Mich-1	na	<200	3281
45	Mich-3	na	<200	3587
45	Mich-4	na	<200	3679
45	M2-1	153.88	562	na
45	M2-3	153.88	486	na
45	M2-2	153.95	458	na
45	M2-4	153.88	284	na
45	R1-1	na	<200	2411
45	R1-2	na	<200	1438
45	R1-3	na	<200	1644

This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.

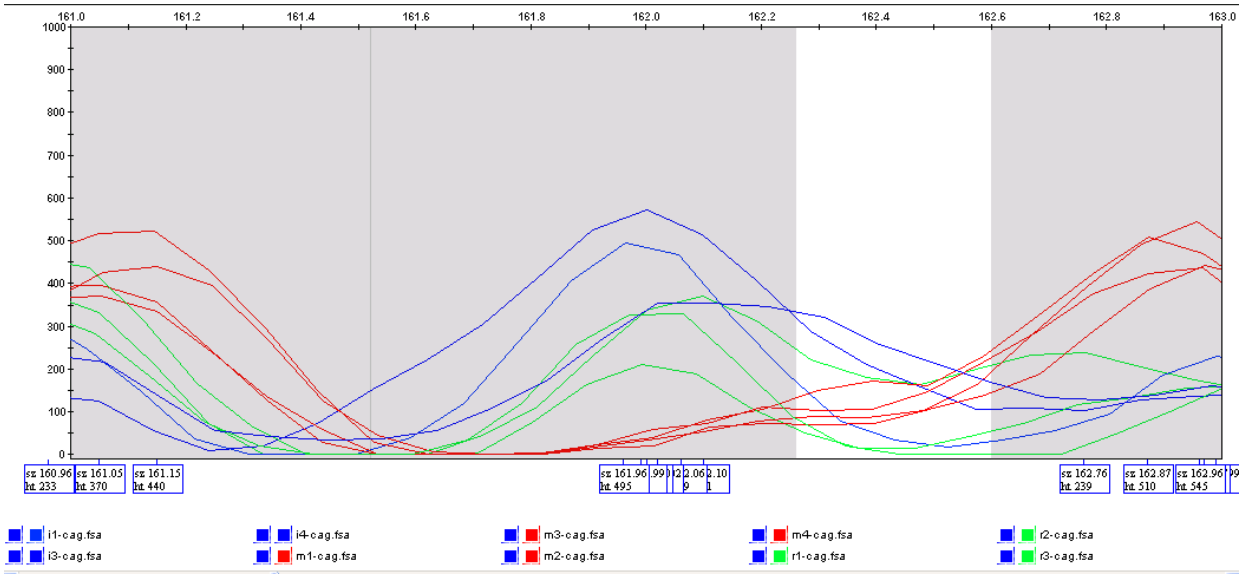

Figure 5. Allele 50 for Primer Pair Mse1-ATC: Eco R1-CAG

In this figure, the grey bin in the center (separated from the one on the left by a dark grey line) demarks Allele 50 at around 162 bp. Organisms from the M2 collection site (marked in red) are shown not to contain the fragment, while organisms from MI and R1 (blue and green) contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CAG.

Table 6. Allele 50 for Primer Pair Mse1-ATC: Eco R1-CAG **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
50	Mich-1	161.96	495	2469
50	Mich-3	162	572	4195
50	Mich-4	162.02	355	3715
50	M2-1	na	<200	na
50	M2-3	na	<200	na
50	M2-2	na	<200	na
50	M2-4	na	<200	na
50	R1-1	162.06	329	1391
50	R1-2	161.99	212	884
50	R3-3	162.1	371	2413

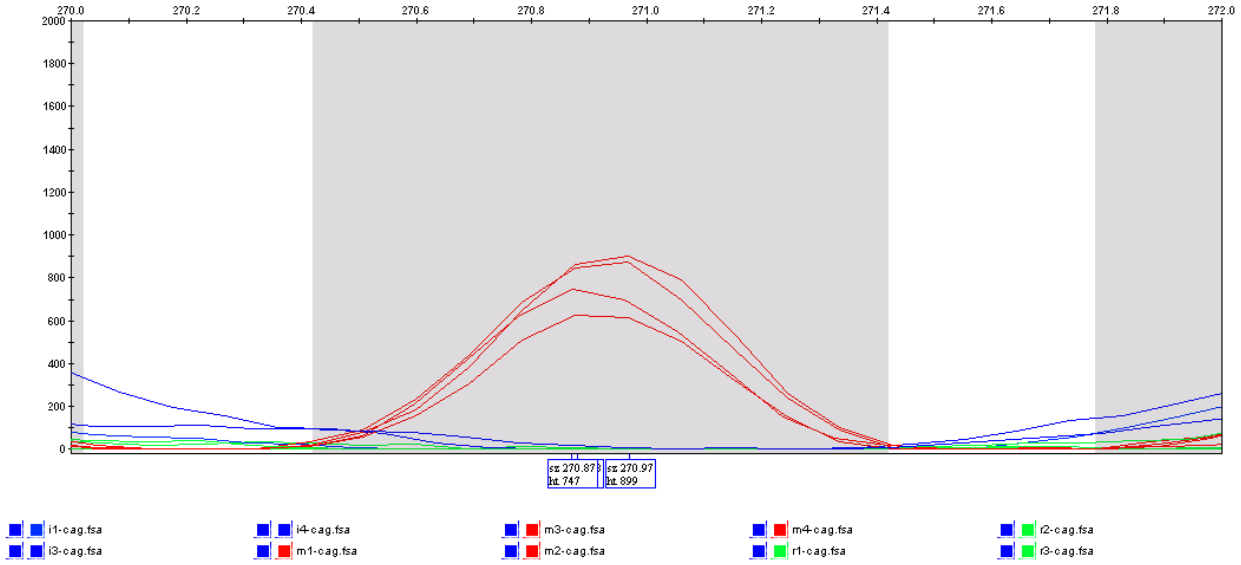

Figure 16. Allele 110 for Primer Pair Mse1-ATC: Eco R1-CAG

In this figure, the grey bin in the center demarks Allele110 at around 271 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CAG.

Table7. Allele 110 for Primer Pair Mse1-ATC: Eco R1-CAG **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
110	Mich-1	na	<200	na
110	Mich-3	na	<200	na
110	Mich-4	na	<20	na
110	M2-1	270.97	875	4724
110	M2-3	270.87	747	3908
110	M2-2	270.97	899	4791
110	M2-4	270.88	624	3306
110	R1-1	na	<200	na
110	R1-2	na	<200	na
110	R1-3	na	<200	na

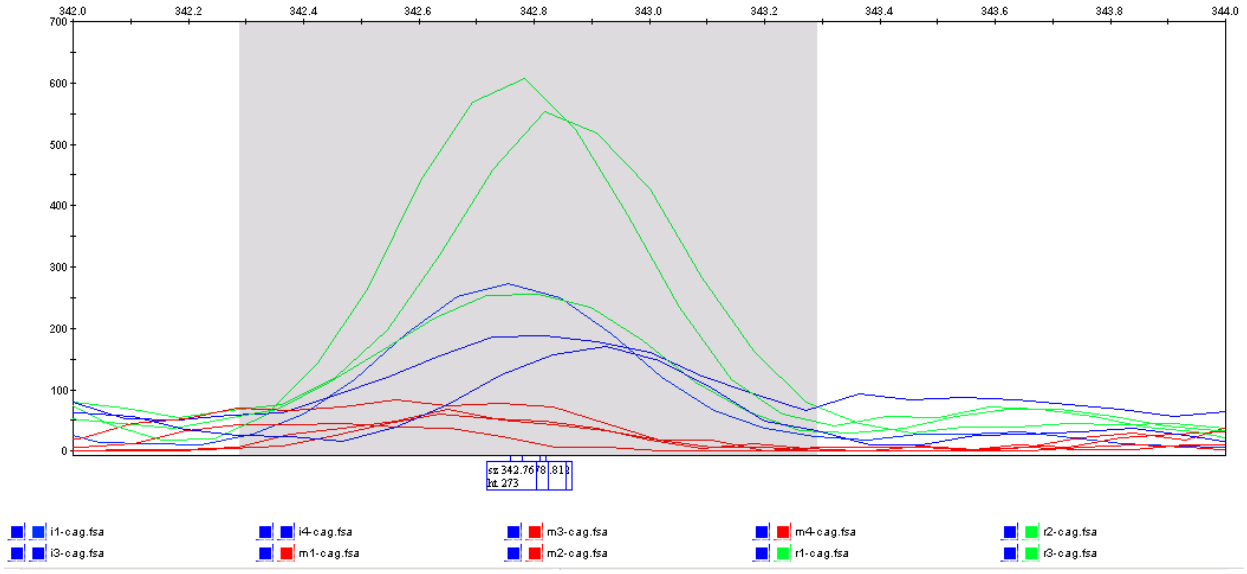

Figure 17. Allele 145 for Primer Pair Mse1-ATC: Eco R1-CAG

In this figure, the grey bin demarks Allele 145 at around 342 bp. Organisms from the M2 collection site (marked in red) are shown not to contain the fragment, while organisms from MI and R1 (blue and green) do contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CAG.

Table 8. Allele 145 for Primer Pair Mse1-ATC: Eco R1-CAG **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
145	Mich-1	342.76	273	1686
145	Mich-3	342.78	190	na
145	Mich-4	342.7	180	na
145	M2-1	na	<200	na
145	M2-3	na	<200	na
145	M2-2	na	<200	na
145	M2-4	na	<200	na
145	R1-1	342.78	608	3611
145	R1-2	342.81	255	1957
145	R1-3	342.82	553	3439

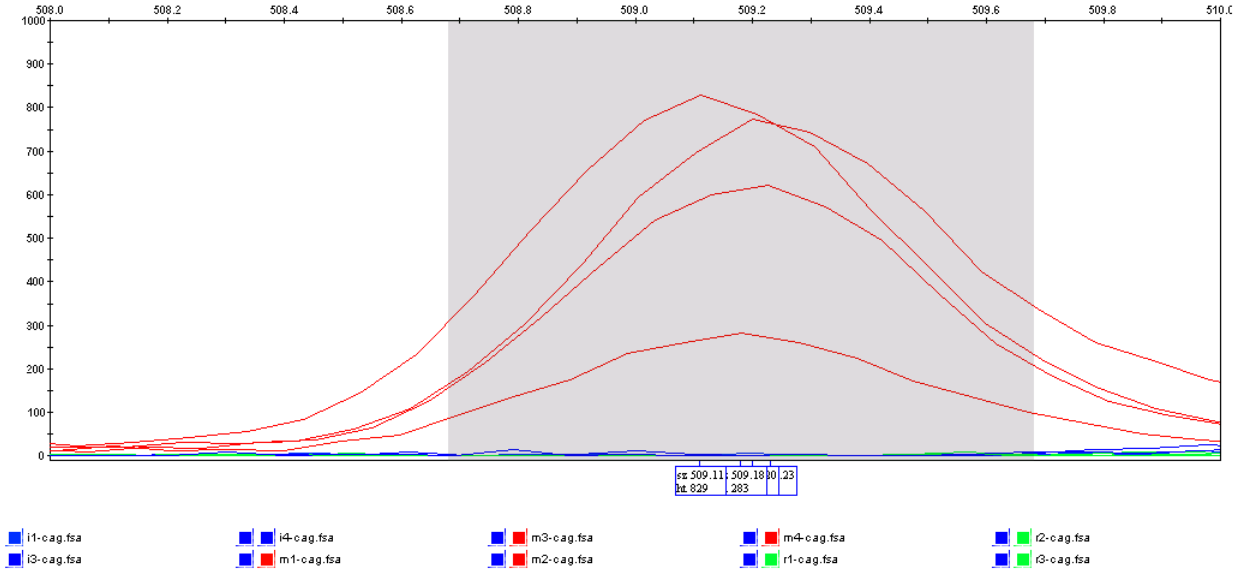

Figure 18. Allele 171 for Primer Pair Mse1-ATC: Eco R1-CAG

In this figure, the grey bin demarks Allele 171 at around 509 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CAG.

Table 9. Allele 171 for Primer Pair Mse1-ATC: Eco R1-CAG **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
171	Mich-1	na	<200	na
171	Mich-3	na	<200	na
171	Mich-4	na	<200	na
171	M2-1	509.18	283	2472
171	M2-3	509.11	829	7383
171	M2-2	509.23	622	5381
171	M2-4	509.2	775	7180
171	R1-1	na	<200	na
171	R1-2	na	<200	na
171	R1-3	na	<200	na

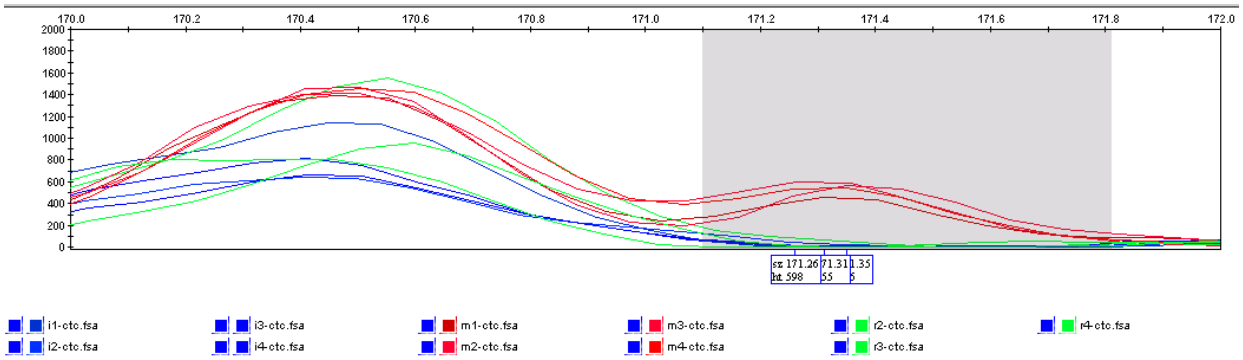

Figure 19. Allele 57 for Primer Pair Mse1-ATC: Eco R1-CTC

In this figure, the grey bin demarks Allele 57 at around 171 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CTC.

Table 10. Allele 57 for Primer Pair Mse1-ATC: Eco R1-CTC **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
57	Mich-1	na	<200	na
57	Mich-2	na	<200	na
57	Mich-3	na	<200	na
57	Mich-4	na	<200	na
57	M2-1	171.31	455	2776
57	M2-2	171.26	598	3231
57	M2-3	171.35	566	3179
57	M2-4	171.25	590	na
57	R1-2	na	<200	na
57	R1-3	na	<200	na
57	R1-4	na	<200	na

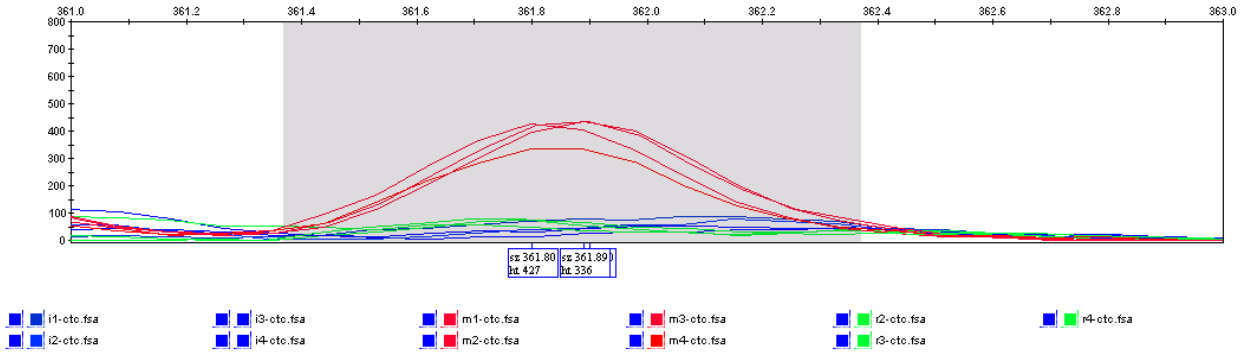

Figure 20. Allele 119 for Primer Pair Mse1-ATC: Eco R1-CTC

In this figure, the grey bin demarks Allele 119 at around 362 bp. Organisms from the M2 collection site (marked in red) are shown to contain the fragment, while organisms from MI and R1 (blue and green) do not contain the fragment. The primer set used to create this PCR product was unlabeled MseI 1-ATC and 6-FAM labeled Eco R1-CTC.

Table11. Allele 119 for Primer Pair Mse1-ATC: Eco R1-CTC **This table elaborates in the preceding figure by listing precise fragment sizes, peak heights, and peak areas.**

Allele designation	Organism	Fragment Size (bp)	height (RFU)	peak area
119	Mich-1	na	<200	na
119	Mich-2	na	<200	na
119	Mich-3	na	<200	na
119	Mich-4	na	<200	na
119	M2-1	361.89	436	2714
119	M2-2	361.9	434	2636
119	M2-3	361.8	427	2733
119	M2-4	361.89	336	2311
119	R1-2	na	<200	na
119	R1-3	na	<200	na
119	R1-4	na	<200	na