April 2013

Crayfish Paternity Analysis Using AFLPs

Sabrina Ann Holzknecht
Worcester Polytechnic Institute

Follow this and additional works at: https://digitalcommons.wpi.edu/mqp-all

Repository Citation
Crayfish Paternity Analysis Using AFLPs

Major Qualifying Project Report

Submitted to the Faculty

of the

Worcester Polytechnic Institute

in partial fulfillment of the requirements for the

Degree of Bachelor of Science

Biochemistry

By

______________________
Sabrina Holzknecht

Advisor:                                                                                                      Advisor:                                                                        
Michael Buckholt                                                                                         Lauren Mathews

April 24, 2013
Abstract

Amplified fragment length polymorphisms are a PCR-based genetic tool used to mark the presence or absence of alleles in a DNA sample. This tool has typically been used for genetic analysis in plants, but research is now being done to see if it is a reliable source for paternity analysis in animals. Samples of *Orconectes limosus* were collected from Quinebaug River during a previous study and mated to produce offspring. A blind AFLP analysis was conducted to see if juveniles from known mothers could be matched to their appropriate fathers. The correct father was determined in two of five families.
Acknowledgements

This project was made possible with the help of many others. I would like to express my gratitude to my advisors, Professors Lauren Mathews and Michael Buckholt, for the assistance throughout this project. I would also like to thank a fellow WPI student, Oliver Hammond, who wrote the analysis program for this MQP.
Contents

Abstract ................................................................................................................................................. 1
Acknowledgements ................................................................................................................................. 2
Figures ................................................................................................................................................... Error! Bookmark not defined.
Tables .................................................................................................................................................. 4
Introduction ........................................................................................................................................... 5
   A. DNA Analysis Methods .................................................................................................................. 6
      1. Molecular Marker Techniques .................................................................................................... 6
      2. DNA Profiling and AFLP .......................................................................................................... 7
   B. Background on Crayfish Characteristics ...................................................................................... 11
Methodology ........................................................................................................................................ 13
   A. DNA Extraction ............................................................................................................................ 13
   B. AFLP Protocol .............................................................................................................................. 14
   C. Sample Preparation ....................................................................................................................... 18
   D. Scoring Bands ............................................................................................................................... 18
   E. Analyzing Allele Tables ............................................................................................................... 21
Results .................................................................................................................................................. 24
Discussion .......................................................................................................................................... 27
Works Cited ......................................................................................................................................... 30
Appendix B: Table of Crayfish Samples Studied ................................................................................ 33
Appendix B: Code for Paternity Analysis Program ............................................................................ 35
Figures
Figure 1 - Gel Image: Representative DNA Extraction .......................................................... 14
Figure 2 - Generation of AFLP Markers (Mueller, 1999) ................................................................. 15
Figure 3 - Additional Settings for AFLP Protocol in Gene Marker V2.40 Software ....................... 19
Figure 4 - Chromatograph for Sample 10F .................................................................................. 21

Tables
Table 1 - Adaptor and Primer Sequences. Modified From the “Wolf lab” Protocol .......................... 17
Table 2 - Portion of Allele Table generated in excel. The table shows the dye color for each primer set (row 1), the called base pair size for each allele bin (row 2), and a presence (1) or absence (0) of that allele for each sample ........................................................................................................ 23
Table 3 - Matched Fathers ............................................................................................................. 24
Table 4 - Matched and Actual Fathers .......................................................................................... 26
Introduction

The field of molecular ecology has been rapidly moving forward with the development and refinement of genetic analysis over the years. The limits of molecular ecology have often been defined by the scope and availability of useful genetic markers. One of the most frequently used techniques in molecular ecological analysis has been microsatellite testing. Microsatellites are repeating sequences of base pairs that allow for a co-dominant loci system—meaning that it can determine whether an individual received loci from one of both parents (Bennett, 2000). Analysis techniques such as this one however, need to have pre-developed markers for the species that is being tested. This development is expensive and time consuming, whereas using a technique that utilizes randomly amplified DNA fragments is more cost effective and easier to generate markers with (Gerber, 2000). Amplified fragment length polymorphism (AFLP) is a molecular genetic analysis technique that uses randomly amplified DNA, and thus is cheaper and easier to begin this analysis than techniques that use pre-determined markers. This technique offers a dominant loci system—meaning that it will only score the absence or presence of loci, not whether they came from one or both parents. While it is mostly used in plant ecology, in recent years AFLP testing has been shown to have potential for use in animal studies (Hoffman, 2012; Huang, 2007; Negrini, 2007). For this blind study, a crayfish species, *Orconectes limosus*, collected from the Quinebaug River in Massachusetts was used to determine the usefulness of AFLP testing in animal genetics. These crayfish were mated to produce juveniles that were analyzed alongside their mothers and potential fathers in order to determine if AFLPs were reliable enough to determine the paternity. In this study, the mothers of the juveniles were known and the fathers were determined using AFLP.
A. DNA Analysis Methods

There are a few effective techniques that are currently used for paternity testing in animals. The most commonly used method is microsatellite analysis. However, Amplified Fragment Length Polymorphisms (AFLPs) have been shown to be a lower cost, quicker technique (Hoffman, 2012). This technique has only been applied to a few animal studies, mostly being utilized in ecological studies. In this section I discuss relevant background on microsatellites, AFLPs and a few other DNA fingerprinting techniques. Also, the effectiveness of AFLPs for other evolutionary biology practices, such as population genetics and hybrid identification, will be examined.

1. Molecular Marker Techniques

DNA fingerprinting refers to a suite of techniques used to identify and compare individuals genetically. This technique uses select primer sites to break DNA into fragments of varying sizes. These fragments are then amplified using a technique called Polymerase Chain Reaction (PCR). These different sized fragments can be detected using electrophoresis (Belkum, 1994). Fingerprinting uses many a number of different techniques, including those such as variable number of tandem repeats (VNTR), random amplification of polymorphic DNA (RAPD), and restriction fragment length polymorphism (RFLP).

Single nucleotide polymorphisms (SNPs) are frequently used in DNA fingerprinting. This technique is similar to microsatellites in that it is also uses co-dominant markers and produces few loci. SNPs are single nucleotide variations on the genome that differ between members of a species (Brookes, 1999). These polymorphisms have a low mutation rate and can be detected in short fragments of less than 150 base pair and so are useful in analysis of degraded samples (Costa, 2008). However, like microsatellite sequences, since specific markers are
required and few loci are generated, if prior sequence information is not available the proper markers cannot be chosen.

Besides AFLPs, there are other more commonly used methods of paternity testing in animals. One of these methods is microsatellite DNA sequencing, which involves the analysis of short tandem repeats (STRs) of base pairs in nuclear DNA. Microsatellites are found in both prokaryotes and eukaryotes and can vary in size from a few tens of bases to about one hundred, and so are small enough to be amplified using PCR (Bennett, 2000). These markers are co-dominant and spread across the euchromatic part of the genome, which has a high gene concentration. Co-dominant makers make this technique more powerful in population genetics where it is important to estimate allele frequencies. In species with weak population differentiation, microsatellites are not as useful for individual-based comparison as AFLPs because only a small amount of markers can be generated using microsatellites, not allowing for enough loci to properly analyze the weak differences (Campbell, 2003). Microsatellites also require prior sequence information on a species in order to select the correct markers for paternity testing (Meudt, 2007). While AFLPs are dominant and so are individually less informative, AFLP techniques yield more numerous loci, do not need prior sequence information, and tend to be more cost-effective then microsatellites.

2. DNA Profiling and AFLP

The technique known as amplified fragment length polymorphism (AFLP) analysis was first developed in 1995 as an alternative DNA fingerprinting method for non-model organisms—organisms that do not have or have little prior known reference sequences—although it is also used on model organisms as well (Vos, 1995). This technique is a combination of two other DNA
fingerprinting techniques known as random amplified polymorphic DNA (RAPD) and restriction fragment length polymorphism (RFLP). As in RFLP analysis, restriction enzymes are used to digest genomic DNA and as in RAPD analysis, arbitrary primers are used in polymorphic chain reaction (PCR) to amplify fragments (Bensch 2005). It is unclear exactly why this method has been underused in wild animal studies when it has been shown to have potential for use in this area, but it is possibly because of the current difficulty in loci analysis (Huang, 2007). The past studies done using AFLP have mostly been concerned with economically important plant or bacterial species that are cultivated for human use. In these cases, a quick, low cost, and versatile genetic maker is ideal for a higher production value (Bensch, 2005).

AFLP is a simple PCR-based technique that can amplify uncharacterized DNA fragments by using arbitrarily cut DNA fragments, making this technique versatile in its uses. For this method, a set of pre-determined sequences for the organism being studied is not necessary to analyze the amplified DNA (Meudt, 2007). Unlike microsatellites, AFLPs have a short start-up time and a low-cost for fragment typing, and so numerous loci (>1000) can be studied, compared to the small amount of loci (<50) that can be used for microsatellites. The major disadvantage of AFLP is that it only generates dominant markers. This means that only the presence or absence of DNA fragments can be detected and there is no differentiation between homozygous and heterozygous genotypes (Bensch, 2005). Another problem encountered with this technique is that with a larger number of arbitrarily sized bands, differentiating which bands represent what loci can be unclear. However, this uncertainty can be minimized by applying protocols to detect these different fragments, such as using fluorescent labeled primers (Hoffman, 2012).

Since AFLP can generate a large number of polymorphic markers, it has been shown to be a reliable method for paternity analysis in ecological studies, although it has been used little in
wild animal studies. AFLPs have a predominately nuclear origin, which has been found to be a useful alternative for markers derived from chloroplast and mitochondrial genomes. These genomes sometimes do not have highly variable alleles and tandem repeats, and therefore are limited in their usefulness for microsatellite animal studies (Meudt, 2007). However, because microsatellites have already become an established technique for animal paternity analysis, AFLPs are often overlooked as an easy and cost effective alternative to the otherwise typically high cost and difficulty of isolating nuclear markers associated with microsatellites (Hoffman, 2012). A very small number of animal paternity tests have been performed on wild animals using AFLP, but as more studies are done this method seems to be reliable for many research questions in which dominant markers can be utilized (Mueller, 1999).

For example, in a study by Huang et al. (2007), duck paternity was tested using AFLPs to determine its accuracy and reproducibility. In this study, they found the peaks to be highly reproducible and that each of the loci in the offspring was derived from one parent or the other (Huang et al, 2007). Negrini et al. (2007) used AFLPs to study the genetic diversity of European cattle in order to determine the genetic distance between the cattle that were first introduced into Europe, and the present species of cattle. They used both microsatellite testing and AFLPs to calculate the genetic distances and found the distances they got using each correlated well. The AFLP data revealed similarities between a type of Asian cattle and the European cattle that the microsatellites did not, meaning that by analyzing a greater number of loci, they were able to discover a new correlation they had previously missed (Negrini et al, 2007). Cameron et al. (2003) reported research on separation of pig lines over the course of seven generations. Nine of these lines raised from the same starting population were analyzed using AFLPs. Within the individual animal lines, the marker scores were constant between the animals, but did not overlap
with other lines. In this case, AFLP also proved to be an effective method of DNA analysis, used here to determine a good selection strategy for increased lean growth rate of the pigs (Cameron, 2003).

There are many different questions within molecular ecology, most of which can be researched using AFLPs. The most often use for AFLP markers is in identifying hybrid loci and backcrosses (Grzebelus, 2001). Hybrids occur when offspring are from two genetically different parents and carry two different alleles of the same gene—one from each parent. In these studies, knowing the allele frequency, the proportion of all copies of a gene that are made up of a particular loci variant, is important to determining how genetically different siblings are (Grzebelus, 2001). AFLPs are useful for these identifications because the technique can generate many markers in a short time. Microsatellites do not produce as large a variety of markers, so hybrids and backcrosses can be analyzed better using AFLP markers (Bensch, 2005). Also, since AFLPs only analyze the allele presence or absence, it is easier to calculate the genetic distance—how far two individuals have diverged from each other—of the hybrids (Grzebelus, 2001).

Population genetics is used to quantify genetic differentiation between individuals of a population as well as to understand genetic difference between populations. AFLP analysis can be useful in these studies because the markers offer a direct comparison between individuals by using a larger number of loci (Campbell, 2003). Typically, microsatellites are used in these studies because they are a highly established method that works for high population diversity because of their co-dominant system. However, because of the low number of loci produced, microsatellites show poor statistical power for weak differentiation. Since AFLPs use randomly generated fragments, the populations can be compared without need of prior sequence information and more loci can be produced. According to Campbell (2003), with more loci there
is a greater chance to find differences in populations with weaker differentiations (Campbell, 2003).

This blinded study was conducted to determine if AFLP is a useful and valid method for the paternity testing of crayfish. In this study, the maternal parent of the juveniles was known, and the AFLP technique was utilized in a blind test in order to determine the paternal parent. This matched paternal parent will then be compared to the actual father to determine the accuracy of the technique. Since AFLPs are a cost effective and quick method of DNA analysis, it would be a good substitute for microsatellite analyses if it can be proven as a reliable method of paternity testing.

B. Background on Crayfish Characteristics

The crayfish species *Orconectes limosus* belongs to the family Cambaridae. This type of crayfish is indigenous to North America. This species, also known as spiny-cheek crayfish, is a smaller crayfish species with an adult total length of 50-54mm and a lifespan of three to four years (Holdich & Black, 2007). They tend to live in larger streams of the more temperate water of the north-eastern USA, also being found in small brooks or rivulets in Canada (Buric et al, 2009). Male and female spiny-cheek crayfish molt twice a year, in May or June, and in August. Females of this species undergo what is called cyclic dimorphism, meaning that they alternate between sexually active and sexually inactive stages (Buric et al, 2009).

*O. limosus* has two mating periods, one in spring and one in fall, followed by spawning period and a short incubation period (Buric et al, 2009). During the mating season females store the male’s sperm in what is called a seminal receptacle. Males deposit sperm in this receptacle using copulatory stylets, also known as sperm tubes (Holdich & Black, 2007). Females are not necessarily fertilized once they are mated, thus having the ability to store sperm from multiple
males. As a result, during the incubation period, in which the female is fertilized by the sperm she is carrying, a female can be fertilized by sperm from multiple males, leading to the possibility of more than one father for her clutch (Buric et al, July 2009). Eggs are typically laid in April or May and are attached to the female for one to three weeks before being hatched in May or June (Holdich & Black, 2007).
Methodology

A. DNA Extraction

DNA was extracted from crayfish leg muscle of seven adult females and forty-one males previously preserved in 90% ethanol. DNA was also extracted from five whole juveniles from the clutch of each of the seven adult females also preserved in 90% ethanol. These individuals were previously used in a behavioral experiment a year prior to this study and the mothers for all the juveniles were known before the samples were analyzed. For each family of offspring it was assumed that only one male sired all the juveniles, however it was possible for a single male to be the father in multiple families, and that the fathers were in the data set (Mathews, personal communication). Either ethanol preserved (90% v/v) whole juveniles or tissue samples (5-10 mg removed from the crayfish legs), was added to 300 µL of Qiagen Cell Lysis Solution and incubated at 65°C for fifteen minutes. The tissue was then homogenized using a microcentrifuge pestle, and 5 µL of Proteinase K was added to each. Tubes were inverted to mix and incubated at 55°C overnight. After incubation, 100 µL protein precipitate solution was added to each sample and the samples were vortexed on high speed for 20 seconds. This mixture was then run for three minutes in a centrifuge at 16,000xG. The supernatant from these samples was added to a new microcentrifuge tubes that contained 300 µL isopropanol and the new mixture was inverted fifty times, and then centrifuged for five minutes at 16,000xG. Next, the supernatant was discarded and 300 µL of 70% ethanol was added. Then, the mixture was centrifuged again for one minute at 16,000xG. Once again, the supernatant was discarded and the pellet was allowed to air dry. After ten minutes, 20 µL of DNA hydration solution was added to the pellet and vortexed at medium speed for five seconds. The samples were incubated at 65°C for one hour before one
microliter of the samples were run on a 1% agarose gel to assess DNA quantity and quality, as seen in Figure 1.

![Gel Image: Representative DNA Extraction](image)

**Figure 1-Gel Image: Representative DNA Extraction**

Gel Image of DNA extraction taken December 4th, 2012 for male samples 29, 30, 32, 34, 35, 36, 43, 44, 45, 46, 49 and 50, seen in the order in lanes three to fourteen. Lane 17 contains 100µg/ml Lamda DNA for intensity comparison and lane 19 contains a DNA ladder.

### B. AFLP Protocol

Similar to RFLP, in AFLP restriction enzymes are used to digest genomic DNA and then adaptors are ligated to the sticky ends of the cut DNA. By using selective enzymes that cut at specific sequences, digestion and ligation can be done in the same step. The adaptors used have sequences that differ from those of the original strand, preventing the new ligation from being re-digested. After this ligation, pre-selective and selective PCR are done in a similar manner to RFLP methods to amplify a subset of the fragments generated (Meudt, 2007). A first series of primers extends into the unknown part of the fragments by one base pair, in theory reducing the number of amplified fragments by a factor of 4. The second series of primers perform a three-base pair extension, after which the number of fragments is in theory reduced by a factor of 64 (Mueller, 1999). This process can be seen in Figure 2. The generated fragments are observed by
using fluorescent labeling of the second series of primers. These labels are different for each of the secondary primers used so as to allow for differentiation between the fragments of similar length elongated by different selective primer (Vuylsteke, 2007).

Figure 2- Generation of AFLP Markers (Mueller, 1999)

(a) Shows a small amount of DNA, collected from 5-10mg of tissue, digested with MseI and EcoRI enzymes. In the restriction and ligation, (b), AFLP adaptors are ligated to the cut ends of the DNA, forming a new DNA with the adaptor sequence, shown in red, and the restricted ends of the DNA fragment, shown in green and blue. During selective amplification, (c), the ends formed by the adaptors are used as primers for the PCR reactions, which are preformed twice. These primers are extended one randomly chosen nucleotide into the sequence before amplification in the first PCR, and then three nucleotides in the sequence in the second reaction.

An AFLP protocol was used to analyze the genomic DNA extracted from the crayfish samples. This protocol begins with a digestion-ligation reaction that is run by adding 1 µL,
approximately 100µg/ml, of genomic DNA to 10.5 µL of Master Mix 1. Per reaction, Master Mix 1 consisted of:

- 1.25 µL 10x T4 ligase buffer
- 0.5 µL 1 mg/mL BSA
- 1 µL 0.5 M NaCl
- 0.5 µL 50 mM Mse I adaptor
- 0.5 µL 5 mM Eco RI adaptor
- 0.15 µL 10,000 U/mL Mse I enzyme
- 0.15 µL 20,000 U/mL Eco RI enzyme
- 0.05 µL 400,000 U/mL T4 DNA ligase
- 6.45 µL de-ionized water

The samples were then incubated at 37°C for 2 hours in a thermal cycler, and held at 4°C until they could be relocated to a -20°C freezer. These reactions were then diluted 1:10 with deionized water in preparation for the pre-selective PCRs (PS-PCR). Four primer combinations were used in the PS-PCR step—Mse-A and Eco-A (PSA), Mse-A and Eco-C (PSB), Mse-C and Eco-C (PSC), and Mse-C, and Eco-A (PSD). Sequences for pre-selective and selective primers can be found in Table 1. In this step 2.5 µL of the diluted digestion-ligation reaction was added to 12.5 µL of Master Mix 2. Per reaction, Master Mix 2 consisted of:

- 1.5 µL 10x Thermopol I buffer
- 0.75 µL 2.5 mM dNTPs
- 0.4 µL 10 mM Mse primer
- 0.4 µL 10 mM Eco primer
- 0.075 µL 5,000 U/mL Taq polymerase
- 9.4 µL deionized water

The PCR reaction was run on these samples using a program titled “Crayfish” in the thermal cycler. This program uses the following steps:

1) 95°C for 2 minutes
2) 95°C for 30 seconds
3) 48°C for 30 seconds
4) 72°C for 1 minute
5) 40 total cycles of steps 2-4
6) 72°C for 10 minutes
7) 4 °C “forever” (samples were held at 4 °C until they could be transferred to -20°C freezer)

Each PS-PCR reaction was diluted 1:10 with deionized water in preparation for the Selective PCR (S-PCR) step. Six sets of primer combinations were used for the S-PCR—Mse-ATC and Eco-ACG (SA), Mse-ATC and Eco-CTC (SB), Mse-ATC and Eco-CAG (SC), Mse-CTC and Eco-ACG (SD), Mse-CTC and Eco-CTC (SE), and Mse-CTC, and Eco-CAG (SF). The S-PCR reactions were carried out for their corresponding Pre-selective reaction, (SA was carried out for PSA samples, SB and SC for PSB, SD for PSD, and SE and SF for PSC). For each reaction, 2.5 µL of the diluted PS-PCR sample was added to 12.5 µL of Master Mix 3. Master Mix 3 consisted of:

- 1.5 µL 10x Thermopol I buffer
- 0.75 µL 2.5 mM dNTPs
- 0.2 µL 10 mM Mse primer
- 0.2 µL 10 mM 6FAM label Eco-ACG primer
- 0.075 µL 5,000 U/mL Taq polymerase
- 9.8 µL deionized water

The S-PCRs were run in the thermal cycler using the program “Crayfish” as before.

Table 1-Adaptor and Primer Sequences. Modified From the “Wolf lab” Protocol

<table>
<thead>
<tr>
<th>Adaptor/Primer Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MseI F adaptor</td>
<td>5’-GAC GAT GAG TCC TGA G-3’</td>
</tr>
<tr>
<td>MseI R adaptor</td>
<td>5’-TAC TCA GGA CTC AT-3’</td>
</tr>
<tr>
<td>EcoRI F adaptor</td>
<td>5’-CTC GTA GAC TGC GTA CC -3’</td>
</tr>
<tr>
<td>EcoRI R adaptor</td>
<td>5’-AAT TGG TAC GCA GTC TAC -3’</td>
</tr>
<tr>
<td>Mse-A preselective primer</td>
<td>5’-GAT GAG TCC TGA GTA AA-3’</td>
</tr>
<tr>
<td>Mse-C preselective primer</td>
<td>5’-GAT GAG TCC TGA GTA AC-3’</td>
</tr>
<tr>
<td>Eco-A preselective primer</td>
<td>5’-GAT GAG TCC TGA GTA AA-3’</td>
</tr>
<tr>
<td>Eco-C preselective primer</td>
<td>5’-GAT GAG TCC TGA GTA AC-3’</td>
</tr>
<tr>
<td>Eco-ACG selective primer</td>
<td>5’-6FAM-ACTTCGTAACATATATCACG -3’</td>
</tr>
<tr>
<td>Eco-CAG selective primer</td>
<td>5’-VIC-ACTTCGTAACATATTCCAG -3’</td>
</tr>
<tr>
<td>Eco-CTC selective primer</td>
<td>5’-NED-ACTTCGTAACATATTCCCT-3’</td>
</tr>
<tr>
<td>Mse-ATC selective primer</td>
<td>5’-GAT GAG TCC TGA GTA AAT C -3’</td>
</tr>
<tr>
<td>Mse-CTC selective primer</td>
<td>5’-GAT GAG TCC TGA GTA ACT C -3’</td>
</tr>
</tbody>
</table>
These primer sequences were adapted from the “Wolf Lab” AFLP protocol published at http://bioweb.usu.edu/wolf/aflp_protocol.htm. For the three Eco-NNN primers, a fluorescent dye 5’ label was used to allow for automated fragment scoring.

C. Sample Preparation

In order to analyze collected DNA samples, they were prepped and sent to Cornell, where there was a lab capable of automatically scoring the fragments. The Eco-NNN primers used in the selective PCR were all fluorescently labeled at the 5’ end in other to do this analysis. The samples from the S-PCR reaction were diluted 1:10 in preparation for shipping to Cornell University’s Life Sciences Core Laboratories Center. Before shipping, a PCR plate was prepared in which each well contained 0.4 µL of LIZ 600 size standard and 13.1µL of HiDi formamide. Then, 0.5µL of the reaction products from each of the three Mse-ATC, and Mse-CTC primer combinations for the diluted S-PCR samples were added to appropriate wells.

D. Scoring Bands

Using Gene Marker V2.4.0 software, I was able to determine which samples were viable for use in paternity testing. First, all of the Mse-ANN samples were loaded into the program and were run without a panel, under the protocol “AFLP” and size standard “GS600”. No panel was used since panels are used to compare peak results against pre-determined markers for a set and this initial run was used to determine that the size standard was of good quality and that there were indeed peaks present in the samples. The size standard “GS600” was used because it corresponded to the “LIZ600” size standard that was used to measure the size fragments generated between 20 and 600 bps. The additional settings used to run the data are shown in Figure 3. Any samples that came up with a poor size standard quality or lacking allele peaks according to the program were re-run through the AFLP protocol again. The same was done for the Mse-CNN marker samples.
Most of the settings used in this program were default settings used for AFLPs. The allele call was changed to start at 40 base pairs and end at 550 to avoid noise from excess primers, and the minimum intensity was changed to 200 to limit the accidental marking of peaks caused by baseline noise.

Next, I constructed a bin panel for the data by using the panel editor. A bin panel determines what peaks in the chromatogram are marked as alleles by creating a lower and upper bound for the width of a peak and calling all peaks in that bin one allele. In the bin editor you can chose to create a panel using the generated data from running the samples without a panel. This gave a projected bin panel based on automated analysis, which can later be refined by hand through alteration of bin location or bin size. Bins in the MseA data set were altered by hand to eliminate over-lapping of bins and to delete projected bins that either marked stutter peaks—noise peaks that are part of another peak—or raised baseline peaks. This was done for all three dyes associated with the MseA sample combinations. While the bins were altered by hand, this...
was done without reference to the identity of any of the samples, and so would not have biased
the inferences made from these peaks. The collective bin panel was then saved as panel “MseA”
and the sample set was re-run using the same size standard and AFLP protocol, but this time
panel “MseA” was applied as well. This generated a peak table based off of the marked
chromatograph, an example of which can be seen in Figure 4. The program marks called alleles
as either “Reject,” “Check,” or “Pass” Based on the same settings shown in Figure 3. Passed
alleles are clear and most likely true, rejected alleles were those at which noise was marked in an
allele bin, and alleles marked check imply that the program is not sure whether to reject or pass
the called allele, meaning the it is probably an allele, but should be checked manually. In order to
determine whether peaks were properly marked, each sample was then analyzed manually. In
cases where baseline noise was mistaken for an allele peak, the marked peaks was deleted from
the allele set, and where clear peaks where not marked or marked as “Check” they were added by
hand. This same process in which a bin panel was created—the new panel named “MseC”—and
the called alleles were reviewed was repeated for MseC samples. After the samples were run and
reviewed, the allele chart was exported to allele from the Gene Marker program, an example
portion of which can be seen in Table 2.
Figure 4-Chromatograph for Sample 10F

The chromatograph above shows the trace scan for the MseA primers of sample 10F. The peaks are scored alleles of the blue, yellow, and green dyed primers. The corresponding dyes for Eco-ACG, Eco-CAG, and Eco-CTC, were fluorescent dyes 6FAM, VIC, and NED.

E. Analyzing Allele Tables

The allele tables were analyzed using a program designed by another WPI student, Oliver Hammond, for this project. This program was written in Visual Basic as a Macro for excel. To run this program, the paternity analysis macro is applied to the excel table, it will then ask for the cells containing the sample names to be selected, and then ask to highlight the cells containing the allele scores. A small portion of the allele table generated in excel can be seen in Table 2. First, this program compares the females to their corresponding juveniles and marks
any ‘1’s they have in common as a ‘0’ in the juvenile sample. It does this by reading the name of the sample in the format of family number followed by either a ‘J’ or ‘F’ to determine whether it is a female, or a juvenile (example: 1F, 1J). Any numbers or letters after these do not affect the program’s sorting, and any male sample is identified by a number, which does not affect how the male sample is treated, followed by an ‘M’ to identify it as a male sample. The newly generated juvenile allele table, in which the common alleles between the mother and child have been changed to ‘0,’ is then compared to the allele profiles of potential fathers to find the fathers with the highest number of matching alleles with a positive ‘1’ score. The program then produces a table that gives the name of the juvenile, the name of the mother, the percentage match of the highest matched father, and the name of the highest matched father. This program assumes that the father was present in the data set, and so will produce a ‘matched’ father for all juveniles, even if the best matched father has a low percent. For the allele table used in this experiment, certain samples were eliminated from the set based on the number of alleles generated. Any child or father sample that had fewer than fourteen markers was determined to have too few alleles to accurately match paternity. This was determined by totaling the alleles for each sample and observing where there seemed to be a split in the allele numbers and where the percent and precision of father matches was ideal. This number was chosen arbitrarily by re-running the analysis program until the possible fathers for each juvenile was limited to one or two higher quality matches.
Table 2—Portion of Allele Table generated in excel. The table shows the dye color for each primer set (row 1), the called base pair size for each allele bin (row 2), and a presence (1) or absence (0) of that allele for each sample.

<table>
<thead>
<tr>
<th></th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
<th>Yellow1</th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>43</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>45</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>46</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>48</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>49</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>53</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>54</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>56</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>59</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>70</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

A Score of ‘0’ means that there was no peak present for that bin in the sample. A score of ‘1’ means a peak was seen at that point.
Results

Using six different primer combinations, 226 different loci were found for the 89 samples tested. A potential paternity chart was generated in excel in which each offspring was matched with the most similar father (Table 3). For each family of juveniles, the father matched for to the most juveniles was determined to be the “matched” father for that family. In case of family fourteen, in which only two juvenile samples were used and both where matched to different fathers, the “matched” father was determined by the father who matched the offspring with the highest percentage of alleles in common. For this study, it was assumed that there was only one father per clutch of children. The percent chance of each family’s “matched” father was determined by taking the average of the percent matches for the juveniles that were matched to that father.

Table 3-Matched Fathers

<table>
<thead>
<tr>
<th>Offspring ID</th>
<th>Mother ID</th>
<th>Father Match</th>
<th>Father ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1J1</td>
<td>1F</td>
<td>0.774336</td>
<td>23M or 8M</td>
</tr>
<tr>
<td>1J2</td>
<td>1F</td>
<td>0.929204</td>
<td>11M</td>
</tr>
<tr>
<td>1J3</td>
<td>1F</td>
<td>0.920354</td>
<td>11M</td>
</tr>
<tr>
<td>1J4</td>
<td>1F</td>
<td>0.958702</td>
<td>11M</td>
</tr>
<tr>
<td>1J6</td>
<td>1F</td>
<td>0.958702</td>
<td>8M</td>
</tr>
<tr>
<td>1J7</td>
<td>1F</td>
<td>0.961652</td>
<td>8M</td>
</tr>
<tr>
<td>1J9</td>
<td>1F</td>
<td>0.961652</td>
<td>8M</td>
</tr>
<tr>
<td>1J10</td>
<td>1F</td>
<td>0.958702</td>
<td>8M</td>
</tr>
<tr>
<td>3J1</td>
<td>3F</td>
<td>0.920354</td>
<td>11M</td>
</tr>
<tr>
<td>3J2</td>
<td>3F</td>
<td>0.889381</td>
<td>11M</td>
</tr>
<tr>
<td>3J3</td>
<td>3F</td>
<td>0.938053</td>
<td>11M</td>
</tr>
<tr>
<td>3J5</td>
<td>3F</td>
<td>0.89823</td>
<td>11M</td>
</tr>
<tr>
<td>5J1</td>
<td>5F</td>
<td>0.924779</td>
<td>11M</td>
</tr>
<tr>
<td>5J2</td>
<td>5F</td>
<td>0.889381</td>
<td>11M</td>
</tr>
<tr>
<td>5J3</td>
<td>5F</td>
<td>0.867257</td>
<td>11M</td>
</tr>
<tr>
<td>5J4</td>
<td>5F</td>
<td>0.942478</td>
<td>11M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Offspring ID</th>
<th>Mother ID</th>
<th>Father Match</th>
<th>Father ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>10J1</td>
<td>10F</td>
<td>0.837758</td>
<td>11M</td>
</tr>
<tr>
<td>10J2</td>
<td>10F</td>
<td>0.917404</td>
<td>8M</td>
</tr>
<tr>
<td>10J3</td>
<td>10F</td>
<td>0.941003</td>
<td>11M</td>
</tr>
<tr>
<td>10J4</td>
<td>10F</td>
<td>0.887906</td>
<td>55M or 11M</td>
</tr>
<tr>
<td>10J5</td>
<td>10F</td>
<td>0.908555</td>
<td>11M</td>
</tr>
<tr>
<td>12J1</td>
<td>12F</td>
<td>0.814159</td>
<td>25M</td>
</tr>
<tr>
<td>12J2</td>
<td>12F</td>
<td>0.853982</td>
<td>8M</td>
</tr>
<tr>
<td>12J3</td>
<td>12F</td>
<td>0.871681</td>
<td>8M</td>
</tr>
<tr>
<td>12J4</td>
<td>12F</td>
<td>0.840708</td>
<td>8M</td>
</tr>
<tr>
<td>12J5</td>
<td>12F</td>
<td>0.831858</td>
<td>25M</td>
</tr>
<tr>
<td>14J1</td>
<td>14F</td>
<td>0.902655</td>
<td>55M</td>
</tr>
<tr>
<td>14J2</td>
<td>14F</td>
<td>0.935103</td>
<td>8M</td>
</tr>
<tr>
<td>17J1</td>
<td>17F</td>
<td>0.902655</td>
<td>55M</td>
</tr>
<tr>
<td>17J2</td>
<td>17F</td>
<td>0.933628</td>
<td>55M</td>
</tr>
<tr>
<td>17J3</td>
<td>17F</td>
<td>0.929204</td>
<td>55M</td>
</tr>
<tr>
<td>17J4</td>
<td>17F</td>
<td>0.929204</td>
<td>55M</td>
</tr>
</tbody>
</table>
For the seven families analyzed, the most likely father for each was determined based on this list of potential fathers. For family one, five out of the eight juveniles tested were best matched to male eight. The average percent of the loci consistent between male eight and the juveniles was determined to be 92%, although it is worth noting that for this sample set four of the juveniles had a 96% match to male eight and one only had a 77% match. For family three, three out of four of the samples were matched to male eleven with a 92% average match of loci. Family five was found to be fathered by male eleven for all four samples, with an average match of 91%. For family number ten, four out of five of the juvenile samples were matched to male eleven with an average match of 90%. Family twelve had five analyzed samples, three of which had male eight as the matched father, with an average percent of 85%. The two juvenile samples analyzed for family fourteen were match to male fifty-five with a 90% loci match and male eight with a 94% chance. Family seventeen had three analyzed juveniles, all who were matched to male fifty-five with a 92% loci match.
Once these results were obtained, they were compared with the list of potential fathers for each family; a side by side comparison can be seen in Table 4. Potential fathers were considered to be any male that the mother had mated with at least once in previous experiments. The potential fathers for family one were males thirteen or eight, for ten it was male eleven, for twelve the father was male fifty, for fourteen the father was twenty-three, and for family seventeen the father as male sixty-six or forty-eight. In comparing the actual fathers to the matched fathers, two out of the seven families were correctly matched. Family one and family ten were determined by the analysis to have to correct father. However it is important to note in these results that male samples thirteen, fifty, sixty-six, and forty-eight were not in the analyzed set of potential fathers.

Table 4-Matched and Actual Fathers

<table>
<thead>
<tr>
<th>Mother</th>
<th>Matched Father Percent Match</th>
<th>Matched Father</th>
<th>Actual Father</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>92%</td>
<td>8</td>
<td>8 or 13*</td>
</tr>
<tr>
<td>3</td>
<td>92%</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>91%</td>
<td>11</td>
<td>8 or 13*</td>
</tr>
<tr>
<td>10</td>
<td>90%</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>12</td>
<td>85%</td>
<td>25</td>
<td>50*</td>
</tr>
<tr>
<td>14</td>
<td>94% or 90%</td>
<td>8 or 55</td>
<td>23</td>
</tr>
<tr>
<td>17</td>
<td>92%</td>
<td>55</td>
<td>66* or 48*</td>
</tr>
</tbody>
</table>

The matched and actual fathers are shown here. Male samples marked with (*) were not analyzed in the data set due to either a low allele count or absence from original 41 male samples.
**Discussion**

By use of AFLP, markers useable for the comparison of female, male, and offspring genomes were generated for paternity analysis of the crayfish *O. limosus*. In this study, two of the seven families analyzed were matched correctly to the actual father; however a few of the actual fathers were not present in the data set that the juveniles were compared to. Table 4 shows the families with their matched and actual fathers, and well as marks these missing fathers. For families twelve and seventeen, none of the correct fathers were present in the set, and for samples one and five, male thirteen was absent from the set. Males fifty and sixty-six had allele counts that were determined to be too low for a reliable match, and therefore were excluded from the data before the paternity analysis was run. The other two male samples, thirteen and forty-eight, were not in the set of original muscle samples provided. The most probable cause of this would be that these specimens died during the prior study before a sample could be taken from them and preserved (Mathews, personal communication). Taking these sample absences into account, families five, twelve, and seventeen can be disregarded when determining the success rate of this experiment. Out of the possible four families that could be correctly matched, two were matched correctly, both with over a 90% match of paternity, giving this experiment a 50% success rate. This success rate is a good starting point to further studies, implying that with refinement, AFLP techniques may be useful for this form of genetic testing in animals.

During the analysis of these crayfish samples, problems were encountered with matching fathers due to the absence of fathers from the data set. One of the reasons for these absences was that the amounts of alleles for some of the samples were determined to be too low, anything under fourteen alleles, to be reliable for determining paternity. Mothers with low allele counts were not eliminated from the dataset since they were still needed for comparison against their
children. Whereas not all of the alleles from these female samples were amplified, the few that were could still be eliminated for the data set of the juveniles for a slightly more accurate loci match to the fathers. In order to limit this problem, an adequate number of loci for the samples to be feasible should be determined before the study and any sample that does not meet that requirement should have been put through the AFLP protocol again. Unfortunately, due to time constraints and limited information about how many loci to expect for each sample, this was not done in this experiment. One way to avoid low allele counts before analysis is frequent gel runs between procedural steps to make sure a large amount of fragments are being recovered. A frequent problem in genotyping leading to this is a low DNA quantity or quality, resulting from extreme dilution of DNA or degradation (Pompanon et al, 2005). By frequently running gels, contamination, extreme dilution and degradation can be observed early on. In order to limit the number of miss-scored alleles after analysis, a mean peak height for each locus in a given bin may be found in order to determine an appropriate threshold for a reliable score (Whitlock et al, 2008). By doing this instead of determining uniform minimum peak intensity for all bins, the error for scored alleles can be minimized since the heights are being compared directly and not generally.

This experiment was also conducted under the assumption that all fathers for the families would be in the muscle sample set of forty-one, which was not the case due to prior circumstances. As a result of this assumption, the paternity analysis program used in this experiment was designed to assume that one of the males in the set analyzed had to be the father, therefore giving a best-matched father for each juvenile. For further studies, the program should be altered to incorporate the possibly that none of males included in the analysis was the father of a particular offspring. As well, in further blind studies used to determine the success rate of
the AFLP technique, any female and juvenile samples for which DNA samples have not been collected from the appropriate male should be excluded from the experiment before the study is conducted. As well, future studies should be run as a comparison between AFLP procedures and another more popular technique such as microsatellite testing in order to establish a more direct relative cost, time efficiency, and result reliability between the two.

The 50% success rate of this experiment is a good indication that, with further refinement, this AFLP technique could be used as a suitable replacement for microsatellites in animal paternity testing. Switching to this technique can make population and paternity studies cheaper and easier to conduct. With the larger loci set generated by AFLP analysis, it is possible that these tests will be more accurate as well. Further blind studies should be conducted using a larger set of families and with a more concrete and certain mating scheme used before DNA analyses.
Works Cited


### Appendix B: Table of Crayfish Samples Studied

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Alleles Found</th>
<th>Used in Paternity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>10FA.fsa</td>
<td>79</td>
<td>Yes</td>
</tr>
<tr>
<td>10J1A.fsa</td>
<td>75</td>
<td>Yes</td>
</tr>
<tr>
<td>10J2A.fsa</td>
<td>69</td>
<td>Yes</td>
</tr>
<tr>
<td>10J4A.fsa</td>
<td>55</td>
<td>Yes</td>
</tr>
<tr>
<td>12J1A.fsa</td>
<td>54</td>
<td>Yes</td>
</tr>
<tr>
<td>23MA.fsa</td>
<td>54</td>
<td>Yes</td>
</tr>
<tr>
<td>10J5A.fsa</td>
<td>51</td>
<td>Yes</td>
</tr>
<tr>
<td>12FA.fsa</td>
<td>50</td>
<td>Yes</td>
</tr>
<tr>
<td>12J5A.fsa</td>
<td>50</td>
<td>Yes</td>
</tr>
<tr>
<td>12J4A.fsa</td>
<td>48</td>
<td>Yes</td>
</tr>
<tr>
<td>3FA.fsa</td>
<td>46</td>
<td>Yes</td>
</tr>
<tr>
<td>25MA.fsa</td>
<td>46</td>
<td>Yes</td>
</tr>
<tr>
<td>1J1A.fsa</td>
<td>45</td>
<td>Yes</td>
</tr>
<tr>
<td>12J2A.fsa</td>
<td>44</td>
<td>Yes</td>
</tr>
<tr>
<td>12J3A.fsa</td>
<td>35</td>
<td>Yes</td>
</tr>
<tr>
<td>5J3A.fsa</td>
<td>34</td>
<td>Yes</td>
</tr>
<tr>
<td>10MA.fsa</td>
<td>30</td>
<td>Yes</td>
</tr>
<tr>
<td>5J2A.fsa</td>
<td>29</td>
<td>Yes</td>
</tr>
<tr>
<td>17J1A.fsa</td>
<td>28</td>
<td>Yes</td>
</tr>
<tr>
<td>3J5A.fsa</td>
<td>28</td>
<td>Yes</td>
</tr>
<tr>
<td>3J2A.fsa</td>
<td>22</td>
<td>Yes</td>
</tr>
<tr>
<td>14J1A.fsa</td>
<td>21</td>
<td>Yes</td>
</tr>
<tr>
<td>55MA.fsa</td>
<td>21</td>
<td>Yes</td>
</tr>
<tr>
<td>10J3A.fsa</td>
<td>19</td>
<td>Yes</td>
</tr>
<tr>
<td>5J1A.fsa</td>
<td>19</td>
<td>Yes</td>
</tr>
<tr>
<td>3MA.fsa</td>
<td>19</td>
<td>Yes</td>
</tr>
<tr>
<td>17J3A.fsa</td>
<td>18</td>
<td>Yes</td>
</tr>
<tr>
<td>3J1A.fsa</td>
<td>18</td>
<td>Yes</td>
</tr>
<tr>
<td>1J2A.fsa</td>
<td>16</td>
<td>Yes</td>
</tr>
<tr>
<td>11MA.fsa</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>17J4A.fsa</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>3J3A.fsa</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>5J4A.fsa</td>
<td>15</td>
<td>Yes</td>
</tr>
<tr>
<td>1J3A.fsa</td>
<td>14</td>
<td>Yes</td>
</tr>
<tr>
<td>8MA.fsa</td>
<td>14</td>
<td>Yes</td>
</tr>
<tr>
<td>28MA.fsa</td>
<td>14</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Alleles Found</th>
<th>Used in Paternity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>9MA.fsa</td>
<td>10</td>
<td>No</td>
</tr>
<tr>
<td>14J4A.fsa</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>30MA.fsa</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>64MA.fsa</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>7MA.fsa</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>14MA.fsa</td>
<td>9</td>
<td>No</td>
</tr>
<tr>
<td>14J2A.fsa</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>14J3A.fsa</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>5J5A.fsa</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>36MA.fsa</td>
<td>8</td>
<td>No</td>
</tr>
<tr>
<td>17J2A.fsa</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>43MA.fsa</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>58MA.fsa</td>
<td>7</td>
<td>No</td>
</tr>
<tr>
<td>1FA.fsa</td>
<td>6</td>
<td>Yes</td>
</tr>
<tr>
<td>50MA.fsa</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>32MA.fsa</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>34MA.fsa</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>35MA.fsa</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>44MA.fsa</td>
<td>6</td>
<td>No</td>
</tr>
<tr>
<td>14FA.fsa</td>
<td>5</td>
<td>Yes</td>
</tr>
<tr>
<td>17J5A.fsa</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>59MA.fsa</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>60MA.fsa</td>
<td>5</td>
<td>No</td>
</tr>
<tr>
<td>29MA.fsa</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>56MA.fsa</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>18MA.fsa</td>
<td>4</td>
<td>No</td>
</tr>
<tr>
<td>66MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>21MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>45MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>46MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>49MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>65MA.fsa</td>
<td>3</td>
<td>No</td>
</tr>
<tr>
<td>1J10A.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>1J6A.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>1J9A.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>51MA.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>Sample Name</td>
<td>Alleles Found</td>
<td>Used in Paternity Analysis</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>3J4A.fsa</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>12MA.fsa</td>
<td>13</td>
<td>No</td>
</tr>
<tr>
<td>17FA.fsa</td>
<td>12</td>
<td>Yes</td>
</tr>
<tr>
<td>14J5A.fsa</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>1J4A.fsa</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>24MA.fsa</td>
<td>12</td>
<td>No</td>
</tr>
<tr>
<td>17MA.fsa</td>
<td>11</td>
<td>No</td>
</tr>
<tr>
<td>5FA.fsa</td>
<td>10</td>
<td>Yes</td>
</tr>
<tr>
<td>5MA.fsa</td>
<td>10</td>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Alleles Found</th>
<th>Used in Paternity Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>52MA.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>54MA.fsa</td>
<td>2</td>
<td>No</td>
</tr>
<tr>
<td>1J7A.fsa</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>22MA.fsa</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>53MA.fsa</td>
<td>1</td>
<td>No</td>
</tr>
<tr>
<td>1J5A.fsa</td>
<td>0</td>
<td>No</td>
</tr>
<tr>
<td>1J8A.fsa</td>
<td>0</td>
<td>No</td>
</tr>
</tbody>
</table>
Appendix B: Code for Paternity Analysis Program

This code was written in Microsoft Visual Basics for Applications and can be applied as a Macro in Excel. In Microsoft excel, open -> File -> Options -> Customize Ribbon, and then add the developer tab to the ribbon and save. Under the developer tab in your excel workbook, either create a new macro using this code, or open -> File -> Import File, and import a file containing this code.

'To use this code you must include the Microsoft Scripting Runtime library. To do this, Click on the tools menu -> References. Scroll down to Microsoft Scripting Runtime and check the box

Dim dictFathers As New Scripting.Dictionary
Dim dictMothers As New Scripting.Dictionary
Dim dictChildren As New Scripting.Dictionary
Dim strPrimerArray() As String

Private Sub BuildDicts()
    Dim rngNameRange As Range
    Dim rngDataRange As Range
    Dim CurSheet As Worksheet
    Dim i As Integer
    Dim j As Integer
    Dim k As Integer
    Dim strFullName As String
    Dim strFullNameArray() As String
    Dim bytCurAlleles() As Byte
    Dim strMotherName As String
    Dim strChildQualifier As String
    Dim CurChild() As clsChild
    Dim intChildCount As Integer
    intChildCount = 0
    Dim strCurInspect As String

    Set CurSheet = ThisWorkbook.ActiveSheet
    Set rngNameRange = Application.InputBox(Prompt:="Select cell range containing names", Type:=8)
    Set rngDataRange = Application.InputBox(Prompt:="Select cell range containing the allele data", Type:=8)
    For i = 1 To rngNameRange.Rows.Count
        'Get Name
        strFullName = rngNameRange.Cells(i, 1)
        If Not strFullName = "" Then
            'Get Alleles
            ReDim bytCurAlleles(0 To rngDataRange.Columns.Count - 1)
            For k = 1 To rngDataRange.Columns.Count
                bytCurAlleles(k - 1) = rngDataRange.Cells(i, k)
            Next k
            'Find designation
            strFullNameArray = StringToArray(strFullName)
            For j = LBound(strFullNameArray) To UBound(strFullNameArray)

            End Sub

'Appendix B: Code for Paternity Analysis Program

This code was written in Microsoft Visual Basics for Applications and can be applied as a Macro in Excel. In Microsoft excel, open -> File -> Options -> Customize Ribbon, and then add the developer tab to the ribbon and save. Under the developer tab in your excel workbook, either create a new macro using this code, or open -> File -> Import File, and import a file containing this code.

'To use this code you must include the Microsoft Scripting Runtime library. To do this, Click on the tools menu -> References. Scroll down to Microsoft Scripting Runtime and check the box

Dim dictFathers As New Scripting.Dictionary
Dim dictMothers As New Scripting.Dictionary
Dim dictChildren As New Scripting.Dictionary
Dim strPrimerArray() As String

Private Sub BuildDicts()
    Dim rngNameRange As Range
    Dim rngDataRange As Range
    Dim CurSheet As Worksheet
    Dim i As Integer
    Dim j As Integer
    Dim k As Integer
    Dim strFullName As String
    Dim strFullNameArray() As String
    Dim bytCurAlleles() As Byte
    Dim strMotherName As String
    Dim strChildQualifier As String
    Dim CurChild() As clsChild
    Dim intChildCount As Integer
    intChildCount = 0
    Dim strCurInspect As String

    Set CurSheet = ThisWorkbook.ActiveSheet
    Set rngNameRange = Application.InputBox(Prompt:="Select cell range containing names", Type:=8)
    Set rngDataRange = Application.InputBox(Prompt:="Select cell range containing the allele data", Type:=8)
    For i = 1 To rngNameRange.Rows.Count
        'Get Name
        strFullName = rngNameRange.Cells(i, 1)
        If Not strFullName = "" Then
            'Get Alleles
            ReDim bytCurAlleles(0 To rngDataRange.Columns.Count - 1)
            For k = 1 To rngDataRange.Columns.Count
                bytCurAlleles(k - 1) = rngDataRange.Cells(i, k)
            Next k
            'Find designation
            strFullNameArray = StringToArray(strFullName)
            For j = LBound(strFullNameArray) To UBound(strFullNameArray)
strCurInspect = strFullNameArray(j)
If (Not IsNumeric(strCurInspect)) Then
    'First non number - set designation
    Select Case strCurInspect
        Case "M"
            'Father
            Call dictFathers.Add(strFullName, bytCurAlleles)
        Case "J"
            ReDim Preserve CurChild(0 To intChildCount)
            Set CurChild(intChildCount) = New clsChild
            strMotherName = 
            strChildQualifier = 
            'We are at J, so continue looping until we get through the numeric
            For k = j + 1 To UBound(strFullNameArray)
                If (Not IsNumeric(strFullNameArray(k))) Then
                    Exit For
                End If
            Next k
            strChildQualifier = Mid$(strFullName, j + 1, k - j)
            strMotherName = Replace$(strFullName, strChildQualifier, "F")
            CurChild(intChildCount).Mother = strMotherName
            CurChild(intChildCount).Alleles = bytCurAlleles
            Call CurChild(intChildCount).ClearFatherCollection
            Call dictChildren.Add(strFullName, CurChild(intChildCount))
            intChildCount = intChildCount + 1
        Case "F"
            Call dictMothers.Add(strFullName, bytCurAlleles)
        Case Else
            End Select
            Exit For
    End If
Next j
Next i
End Sub

Public Function StringToArray(ByVal strToSplit As String) As String()
    strToSplit = StrConv(strToSplit, vbUnicode)
    StringToArray = Split(strToSplit, Chr(0))
    ReDim Preserve StringToArray(UBound(StringToArray) - 1)
End Function

Private Sub ProcessChildren()
    For Each CurChild In dictChildren.Keys()
        Call FindFather(dictChildren(CurChild))
        Call FindFatherPure(dictChildren(CurChild))
    Next
    'By this point all children should have mothers and fathers
End Sub
Private Sub FindFather(Child As clsChild)
    Dim bytAllelesToFInd() As Byte
    Dim dblHighestChance As Double
    Dim dblCurChance As Double
    Dim strFather As String
    strFather = ""
    dblHighestChance = 0#
    dblCurChance = 0#
    bytAllelesToFInd = ChildAllelesNoMom(Child.Alleles, dictMothers.Item(Child.Mother))
    Child.lngFather = dictFathersReversed(Join(bytAllelesToFInd, "|"))
    Child.ClearFatherCollection
    For Each strFatherName In dictFathers.Keys()
        dblCurChance = GetFatherChance(bytAllelesToFInd, dictFathers(strFatherName))
        If ((dblCurChance > dblHighestChance) And (dblCurChance > 0)) Then
            Child.ClearFatherCollection
            Child.Father.Add (strFatherName)
            dblHighestChance = dblCurChance
        ElseIf ((dblCurChance = dblHighestChance) And (dblCurChance > 0)) Then
            Child.Father.Add (strFatherName)
        End If
    Next
    'Set the final father chance
    Child.FatherChance = dblHighestChance
End Sub

Private Sub FindFatherPure(Child As clsChild)
    Dim bytAllelesToFInd() As Byte
    Dim dblHighestChance As Double
    Dim dblCurChance As Double
    Dim strFather As String
    strFather = ""
    dblHighestChance = 0#
    dblCurChance = 0#
    bytAllelesToFInd = Child.Alleles
    Child.ClearFatherPureCollection
    For Each strFatherName In dictFathers.Keys()
        dblCurChance = GetFatherChance(bytAllelesToFInd, dictFathers(strFatherName))
        If ((dblCurChance > dblHighestChance) And (dblCurChance > 0)) Then
            Child.ClearFatherPureCollection
            Child.FatherPure.Add (strFatherName)
            dblHighestChance = dblCurChance
        ElseIf ((dblCurChance = dblHighestChance) And (dblCurChance > 0)) Then
            Child.FatherPure.Add (strFatherName)
        End If
    Next
    'Set the final father chance
    Child.FatherPureChance = dblHighestChance
End Sub

Private Function ArraysEqual(Array1 As Variant, Array2 As Variant) As Boolean
    ArraysEqual = (Join(Array1, "") = Join(Array2, ""))
End Function

Private Function GetFatherChance(ChildAlleles() As Byte, FatherAlleles() As Byte) As Double
Dim i As Integer
Dim intNumSame As Integer
Dim intNumAlleles As Integer

intNumAlleles = 0
intNumSame = 0

For i = LBound(ChildAlleles) To UBound(ChildAlleles)
    If (ChildAlleles(i) <> 0) Or True Then
        If ChildAlleles(i) = FatherAlleles(i) Then intNumSame = intNumSame + 1
        intNumAlleles = intNumAlleles + 1
    End If
Next i

If intNumAlleles = 0 Then
    GetFatherChance = 0
Else
    GetFatherChance = CDbl(intNumSame) / CDbl(intNumAlleles)
End If

End Function

Private Function ChildAllelesNoMom(ChildAlleles As Variant, MotherAlleles As Variant) As Byte()
Dim ReturnArray() As Byte
ReDim ReturnArray(0 To UBound(ChildAlleles))
For i = LBound(ChildAlleles) To UBound(ChildAlleles)
    'This only works if the child has all of the same alleles is the mother.
    'ReturnArray(i) = ChildAlleles(i) Xor MotherAlleles(i)
    'Below is if for some reason the above doesn't work
    ReturnArray(i) = IIf((ChildAlleles(i) = 1 And MotherAlleles(i) <> 1), 1, 0)
Next i

ChildAllelesNoMom = ReturnArray

End Function

Private Function PrintOutResults()
Dim ResultsSheet As Worksheet
Dim intMaxNumRegFathers As Integer
Dim i As Integer
Dim j As Integer
'Dim strChildrenKeys() As String
'Dim strChildrenKeys = dictChildren.Keys

Set ResultsSheet = ThisWorkbook.Worksheets.Add
If WorksheetExists("Paternity Results") Then
    i = 2
    While WorksheetExists("Paternity Results " & CStr(i))
        i = i + 1
    Wend
    ResultsSheet.Name = "Paternity Results " & CStr(i)
Else
    ResultsSheet.Name = "Paternity Results"
End If
End If

intMaxNumRegFathers = 1

For i = 1 To dictChildren.Count
    If dictChildren.Item(dictChildren.Keys(i - 1)).Father.Count > intMaxNumRegFathers Then intMaxNumRegFathers = dictChildren.Item(dictChildren.Keys(i - 1)).Father.Count
    Next i

ThisWorkbook.Activate
ResultsSheet.Select
With ResultsSheet
    .Cells(1, 1) = "Child Name"
    .Cells(1, 2) = "Mother Name"
    .Cells(1, 3) = "Father Chance"
    .Cells(1, 4) = "Father Name"
    .Cells(1, 4 + intMaxNumRegFathers) = "Father Chance Pure"
    .Cells(1, 4 + intMaxNumRegFathers + 1) = "Father Name Pure"

    For i = 1 To dictChildren.Count
        .Cells(i + 1, 1) = dictChildren.Keys(i - 1)
        .Cells(i + 1, 2) = dictChildren.Item(dictChildren.Keys(i - 1)).Mother
        .Cells(i + 1, 3) = dictChildren.Item(dictChildren.Keys(i - 1)).FatherChance
        For j = 1 To dictChildren.Item(dictChildren.Keys(i - 1)).Father.Count
            .Cells(i + 1, 3 + j) = dictChildren.Item(dictChildren.Keys(i - 1)).Father.Item(j)
        Next j
        .Cells(i + 1, 4 + intMaxNumRegFathers) = dictChildren.Item(dictChildren.Keys(i - 1)).FatherPureChance
        For j = 1 To dictChildren.Item(dictChildren.Keys(i - 1)).FatherPure.Count
            .Cells(i + 1, 4 + intMaxNumRegFathers + j) = dictChildren.Item(dictChildren.Keys(i - 1)).FatherPure.Item(j)
        Next j
    Next i

End With

End Function

Private Function WorksheetExists(WorksheetName As String) As Boolean
    Dim blnRet As Boolean
    blnRet = False
    For Each CurSheet In ThisWorkbook.Worksheets
        If CurSheet.Name = WorksheetName Then
            blnRet = True
            Exit For
        End If
    Next
    WorksheetExists = blnRet
End Function

Private Sub Cleanup()
    'Dim dictFathersReversed As New Scripting.Dictionary
    'Dim dictMothers As New Scripting.Dictionary
    'Dim dictChildren As New Collection
    dictFathers.RemoveAll
    Set dictFathersReversed = Nothing
End Sub
dictMothers.RemoveAll
Set dictMothers = Nothing

dictChildren.RemoveAll
Set dictChildren = Nothing

End Sub

Public Sub ConductAnalysis()
    Call BuildDicts
    Call ProcessChildren
    Call PrintOutResults
    Call Cleanup
End Sub

Public Sub FormatAlleles()
    Call BuildDicts
    Call BuildPrimerArray
    Call PrintOutFormattedAlleles
End Sub

Private Sub BuildPrimerArray()
    Dim rngPrimerRange As Range
    Set rngPrimerRange = Application.InputBox(Prompt:="Select Primer Range", Type:=8)
    ReDim strPrimerArray(0 To rngPrimerRange.Columns.Count - 1)
    For k = 1 To rngPrimerRange.Columns.Count
        strPrimerArray(k - 1) = rngPrimerRange.Cells(1, k) & " " & rngPrimerRange.Cells(2, k)
    Next k
End Sub

Private Function PrintOutFormattedAlleles()
    Dim ResultsSheet As Worksheet
    Dim i As Integer
    Dim j As Integer
    Dim intCurRow As Integer
    Dim intCurCol As Integer
    Dim strCurSubject As String
    Dim bytAlleleArray() As Byte
    intCurRow = 1
    Set ResultsSheet = ThisWorkbook.Worksheets.Add
    If WorksheetExists("Formatted Alleles") Then
        i = 2
        While WorksheetExists("Formatted Alleles " & CStr(i))
            i = i + 1
        Wend
        ResultsSheet.Name = "Formatted Alleles " & CStr(i)
    Else
        ResultsSheet.Name = "Formatted Alleles"
    End If
End Function
End If

ThisWorkbook.Activate
ResultsSheet.Select
With ResultsSheet

.Cells(intCurRow, 1) = "Mothers"
intCurRow = intCurRow + 1
For i = 1 To dictMothers.Count

strCurSubject = dictMothers.Keys(i - 1)
.Cells(intCurRow, 1) = strCurSubject
intCurCol = 2
For j = 1 To UBound(strPrimerArray) + 1
If dictMothers.Item(strCurSubject)(j - 1) = 1 Then
.Cells(intCurRow, intCurCol) = strPrimerArray(j - 1)
intCurCol = intCurCol + 1
End If
Next j
intCurRow = intCurRow + 1
Next i
.Cells(intCurRow, 1) = "Fathers"
intCurRow = intCurRow + 1
For i = 1 To dictFathers.Count

strCurSubject = dictFathers.Keys(i - 1)
.Cells(intCurRow, 1) = strCurSubject
intCurCol = 2
For j = 1 To UBound(strPrimerArray) + 1
If dictFathers.Item(strCurSubject)(j - 1) = 1 Then
.Cells(intCurRow, intCurCol) = strPrimerArray(j - 1)
intCurCol = intCurCol + 1
End If
Next j
intCurRow = intCurRow + 1
Next i
.Cells(intCurRow, 1) = "Children"
intCurRow = intCurRow + 1
For i = 1 To dictChildren.Count

strCurSubject = dictChildren.Keys(i - 1)
.Cells(intCurRow, 1) = strCurSubject
intCurCol = 2
bytAlleleArray = dictChildren.Item[strCurSubject].Alleles
For j = 1 To UBound(strPrimerArray) + 1
If bytAlleleArray(j - 1) = 1 Then
.Cells(intCurRow, intCurCol) = strPrimerArray(j - 1)
intCurCol = intCurCol + 1
End If
Next j
intCurRow = intCurRow + 1
Next i

End With

End Function