Developing Restriction Associated DNA Tags for the Ion Torrent Personal Genome Machine

A Thesis submitted
to the Graduate School
Valdosta State University

in partial fulfillment of requirements
for the degree of

MASTER OF SCIENCE

in Biology

in the Department of Biology
of the College of Arts and Sciences

August 2014

Fiifi Agyabeng-Dadzie

BS, Kwame Nkrumah University of Science and Technology, 2008

This thesis, "Developing Restriction Associated DNA Tags for the Ion Torrent Personal Genome Machine," by Fiifi Agyabeng-Dadzie, is approved by:

Thesis Chair

John F. Elder, Ph.D.
Professor of Biology

Brian C. Ring, Ph.D.
Associate Professor of Biology

Committee Members

David Bechler, Ph.D.
Professor of Biology

Donna Gosnell, Ph.D.
Associate Professor of Chemistry

Interim Dean of the Graduate School

James T. LaPlant, Ph.D.
Professor of Political Science

FAIR USE

DUPLICATION

I authorize the Head of Interlibrary Loan or the Head of Archives at the Odum Library at Valdosta State University to arrange for duplication of this thesis for educational or scholarly purposes when so requested by a library user. The duplication shall be at the user's expense.

Signature 

I refuse permission for this thesis to be duplicated in whole or in part.

Signature _____

ABSTRACT

The identification and classification of similar cryptic species such as *Kryptolebias marmoratus* and *Kryptolebias ocellatus,* as distinct species has always been problematic within the biological sciences. Though *Kryptolebias marmoratus* and *Kryptolebias ocellatus* are both known to be self-fertilizing hermaphrodites, out crossing has been observed resulting in heterozygous offspring. Recently, a viable hybrid (Gitmo) between *K. marmoratus* and *K. ocellatus* has been isolated in the Valdosta State University aquatic lab. This hybrid questions the proposed species status of these fishes.

Classifications of organisms are usually based on both morphological and genotypic studies. Based on morphological studies, both fishes were thought to be synonymous because of very similar anatomical and physiological characteristics. Mitochondrial studies (genotypic analysis) of these fishes have suggested otherwise. Unfortunately, thorough whole genome studies are very expensive and labour intensive. Methods such as Restriction Fragment Length Polymorphisms (RFLP), Amplified Fragment Length Polymorphisms (AFLP) and microsatellite analysis are more commonly used. These methods, though helpful, make analyses based on the length polymorphisms of the DNA fragments and only a small portion of the genome. Furthermore, surveys yield a very limited amount of genetic information.

A more reliable method of analyzing large numbers of these genotypic markers concurrently would be extremely valuable for taxonomic questions. Developing Restriction Associated DNA Tag (RADTag) genome sequencing using methods for the Ion Torrent Personal Genome Machine (PGM) provides a way not only to cut down the cost of genomic sequencing but also provides a method that performs thousands of Single

Nucleotide Polymorphisms (SNP) analyses at once, without the use of a reference sequence. Analysis of RADTag sequences on public domains “Stacks” software displayed thousands of markers present in the Hon9 (*Kryptolebias marmoratus*) genome and confirms the Hon9 strain to be highly homozygous (0.999) across all loci. This achievement paves the way for further genomic studies with high throughputs and accuracy. This RADtag genome will provide the information needed for resolving the questionable relationship between *K. marmoratus* and the presumptive *K. ocellatus* cryptic species.

Key words: RAD-Tags, SNP analysis, AFLP, RFLP, cryptic species

TABLE OF CONTENTS

LIST OF FIGURES

Figures are found in Appendix A

LIST OF TABLES

Tables are found in Appendix B

ACKNOWLEDGEMENTS

I will like to thank God for guiding me through this journey. I cannot express enough gratitude to my committee, especially Prof. J. F. Elder and Associate Prof. B. Ring, for helping me through every step of the way, both academically and socially. I would also like to thank my family and friends, for not giving up on me and always being there to encourage me. To the department of Biology I say thank you for giving me the chance to better myself in this great institution, Valdosta State University.

Chapter I

INTRODUCTION

Cryptic Species

Cryptic species are distinct species erroneously classified to be of the same taxa, based on similarities in morphology (Pfenninger & Schwenk, 2007) and occasionally physiology. The causes of these cryptic natures are still unclear and several reasons such as convergent evolution due to habitat, latitude interspecific interactions, and genetics (de Knijff, 2014) have been suggested. The generally accepted reason is selection resulting in convergent evolution. Behavioral selection and physiological selection of species found in the same habitat and under the similar environmental condition are more likely to exhibit morphological resemblance (Bickford, 2007).

Kryptolebias marmoratus, also known as mangrove killifish, and *Rivulus ocellatus* are examples of presumed cryptic species that belong to the Aplocheilidae family and order Cyprinodontiformes (Figure 1). *Kryptolebias marmoratus* can be found in close association with Red Mangrove Coastal habitats from Florida through Bahamas, Yucatan Peninsula, Venezuela and Guianas as well as Cuba and other Caribbean islands (Costa, 2004a, 2004b). *Kryptolebias ocellatus* is mostly found in the southeastern coastal regions of Brazil (Hensel, 1868). For many years, 1906-1984, these fishes were once thought to be synonymous based on morphological classification (Santos, 1997). The idea of *K. marmoratus* and *K. ocellatus* being conspecific was based on an aquarium bred specimens whose questionable lineage was believed to be from Greta Fundra. By using a

single specimen from Rio de Janeiro for the classification, Hensel (1868) declared *K. ocellatus* was synonymous to *K. marmoratus* and even a third species *K. caudomarginatus*. Not only were they morphologically similar, physiological comparison showed both *K. marmoratus* and *K. ocellatus* to be self-fertilizing androdioecious vertebrates with emersion capabilities (Noerdlic, 2006) due to modified gills and cutaneous respiration abilities (Grizzle & Thiyagarajah 1987).

Genotypic Classification

The development of advanced molecular biological methods such as PCR, microsatellites, DNA sequencing, among other new methods have been very efficient in identifying and classifying organism. Several cryptic species have also been clearly distinguished by these methods even though they were morphologically or physiologically identical (Hebert et al*.,* 2004). In a span of about 2 decades, the amount of molecular research on cryptic species increased tremendously thereby resolving many questions of identity among cryptic species (Figure 2). These methods rely on the identification of few genetic markers that are diagnostic of a particular cryptic species pairs.

Molecular Markers

In population and genomic analysis, the laboratory identification and development of known genetic markers is essential. These markers provide the genetic information for downstream identity and large population studies. Molecular markers also known as genetic markers are any identifiable DNA polymorphisms that are known in the genome of an organism. These markers may also serve as reference points for the analysis of adjacent unknown sections of an organism's DNA, be it nuclear or mitochondrial. The

use of genetic markers has improved the amount and efficiency of research into genetic diversity from the population to individual level (Hillis et al*.,* 1996). A number of techniques relying on different loci types such as microsatellite, Amplified Fragment Length Polymorphisms (AFLP), Single Nucleotide Polymorphisms (SNP) analyses, DNA sequencing, etc., are currently used as genetic markers (Tanya and Kumar, 2010). There are two types of genetic markers, type 1 and type 2, based on their location in the genome. Usually a type 1 marker, e.g. allozymes, are found within coding regions and those found in noncoding regions e.g. microsatellites are known as type 2, (Tanya & Kumar, 2010). Generally, type 2 markers are used in genotypic classification studies due to the relatively low cost of such experiments when compared to sequencing and the presumed evolutionary neutrality of such loci.

Restriction Associated DNA Tags (Radtags)

Microsatellite and AFLP analyses make comparisons among allelic fragment lengths hence, some mutations such as insertions and deletions can determine the outcome of these methods. Such studies therefore provide information of limited type for only small sample of loci types. Therefore, analyzing the DNA nucleotide sequence is a better option providing far more information and resolution to limit the probability of wrongful classification. A comparison of mitochondria gene sequences of *K. marmoratus* and *K. ocellatus* (Tatarenkov et al*.,* 2009) have suggested these fishes to be grouped as sister taxa (Figure 3 and 4). With the development of Next Generation Sequencing (NGS) methods, sequencing of whole genomes has been made easier having much higher output than prior Sanger dideoxynucleotide sequencing. On the down side, NGS machines are

expensive and analyzing the huge outputs of sequence data can become a problem with current softwares.

Restriction associated DNA Tags (RADTags) are DNA sequence markers that are produced by analyzing restriction digested DNA. RADTag sequencing involves the sampling of the genome for homologous locations then identifying SNP markers present at those loci (Catchen et al., 2013). Since RADTag fragments have unique sites that can be easily recognized, comparing and analyzing large data sets of fragments surveyed across whole genomes can be performed providing fine scale, accurate genome surveys. Restriction associated DNA sequence proprietary methods have been developed on the Solexa NGS machine (Illumina, San Diego) using *Sbf 1* restriction enzyme and produced fragment sequences length of approximately 50bp (Amores et al*.,* 2011). The Ion Torrent Personal Genome Machine (PGM) (Life Technologies, Carlsbad) produces approximately 100 Mb, 1 Gb and 2 Gb on the 314, 316 and 318 chip respectively, with an optimized sequenced fragment length not less than 100bp. This implies that more loci can be identified and analyzed. In order to make this unique and efficient mode of sequencing (RADTag sequencing) applicable to cryptic species analysis, more efficient and more practical for less heavily funded labs, this project was performed to develop RADTag libraries and methods for genomes to be sequenced on the Ion Torrent PGM due to the relatively low cost of this machine and its extremely high sequencing efficiency.

Chapter II

MATERIALS AND METHODS

Location

All experiments were performed within the molecular biological laboratories of the Department of Biology, College of Arts and Sciences at Valdosta State University, Valdosta, Georgia from 2012 to 2014.

Ethics Statement

The Animal Use Protocol (AUP-00045-2012) for handling fish material was adhered to (Appendix D) and was approved by the Valdosta State University Institutional Animal Care and Use Committee under Animal Welfare Assurance Number A4578-01. Fish samples were obtained from cultures in the aquatic lab of the Department of Biology. Fish were bred into a new generation before they were selected for euthanizing. Selection was not based on sex or age but on the size of the fish. Only Two Hon 9 (strain designation) (*Kryptolebias marmoratus*) individuals of approximately 35 mm were needed to provide genome samples for this study.

Mitochondria and Nuclei Separation

The Hon 9 specimens were euthanized using standard anesthetic methods prior to pithing. Muscle and soft tissue samples of approximately 300 mg per individual were collected while making sure to exclude the gut and its contents to avoid bacterial DNA contamination. Euthanized fish were washed with a standard 1X PBS wash buffer (Green Sambrok, 2010) in order to remove excess blood and external contaminants. Tissues were

minced on chilled glass plates that had been sterilized with 90% ethanol. In order to prevent over sampling of mitochondrial sequences, mitochondria and nuclei separation was done by means of a modified procedure using chilled Isolation Buffer consisting of 250 mM sucrose, 10mM Tris-HCl pH 7.4 and 0.1 mM EGTA (Chandel Lab). All solutions and instruments were pre-chilled before use and all operations were carried out on ice. Approximately 200 mg of minced sample was placed in a Dounce Homogenizer with 2 mL of Isolation Buffer. Not more than 10 strokes were applied with both the A and B pestle sequentially to homogenize cells and release the nuclei. An initial centrifugation at 1,000 g for 10minutes at 4^0c on a Table Top Centrifuge (Thermo IEC Multi RF) of the homogenate was performed to produce pellets from the larger tissue fragments which contained the nuclei by sucrose gradient. The supernatant was transferred into a new centrifuge tube and centrifuged at 12,500 g for 15-20minutes at 4^0C in order to produce pellets containing the mitochondria. The nuclei and mitochondria pellets were then subjected to DNA extraction procedures separately.

DNA Extraction

Nuclei and mitochondria pellets were each separately re-suspended in equal mass-volume ratios of 2% SDS lyses buffer with 1 μL of proteinase K and incubated at 55^0C overnight. Samples were agitated every 20 minutes within the first hour. A 5 minute centrifugation at 13,000 rpm was performed at the end of the incubation period to pellet cellular debris from the nuclear and mitochondria DNA. DNA was further purified by subsequent extraction with phenol, phenol-chloroform, chloroform-Isoamyl and chloroform. Extractions were followed by centrifugation at 13,000 rpm for 5 minutes. In each case the supernatant was collected while trying not to disturb the inter-phase and

transferred to a new microtube. To produce pellets of DNA, 200 mM NaCl was added, mixed and topped off with approximately 2X volumes of absolute 100% Ethanol, mixed and incubated at -20^{0}C overnight to ensure efficient precipitation of DNA. Samples were then centrifuged at 13,000 rpm for 30 minutes to produce pellets of nucleic acids. Nucleic acid pellets were washed with 70% ethanol to remove the excess salts and air dried to remove residual ethanol. DNA was re-suspended in a solution made up of 100 µL of 10 mM Tris-HCl and 1 mM EDTA pH 8.0 buffers. Quality of the sample and fragment size were analyzed using the Nano drop 2000 ((ThermoScientific, Wilmington) and Agilent 2200 TapeStation System, cat # G2964AA, with Genomic DNA Screen Tape, part # 5067-5365 (Agilent Technologies, Santa Clara) then stored at -20^{o}C.

Library Preparation

Library preparation involves restriction digestion of DNA, ligation of the adaptors containing PCR primer sites and sites necessary for the Ion Torrent sequencing method to the target fragments, size selection for the appropriate fragment size and PCR amplification to produce an adequate number of fragments for sequencing. Three different libraries were built from the same genomic DNA (gDNA) stock samples using three different techniques: technique I (sample A), technique II (sample B) and technique III (sample C) that are described below. These techniques are a hybrid modification of the NEBNext Fast DNA Library Prep Set for Ion Torrent, NEB # E6270S/L, (New England BioLabs, Ipswich) and the RADTag sequencing protocol for Solexa - Illumina (Amores et al*.,* 2011). Products of each step of the modified techniques were analyzed with the Agilent 2200 TapeStation System, cat. # G2964AA, and the D1K Screen Tape

part # 5067-5361(Agilent Technologies, Santa Clara), capable of detecting between 75 – 1000 pg/µL to verify the quality of the DNA before proceeding to the next stage.

Genomic DNA Digestion

DNA samples were treated with RNAseA to remove excess RNA that might be present in the sample. Sample digestion was done using *Pst 1* restriction enzyme, cat. # R3140S, (New England BioLabs, Ipswich), which is a 6 bp cutter, has a GC-rich recognition site (5'-CTGCA^G-3') and cuts approximately 100,000 times in Teleost genomes. A 50 µL restriction digest comprised of 1 µL *Pst 1*enzyme, 5 µL of Cutsmart buffer and 1µg of gDNA was incubated at 37^{o}C for 1 hour in technique I and 2 hours in technique III. The extended length of the incubation period for technique III was to produce shorter fragment lengths. The restriction enzyme was subsequently deactivated at 80^{o}C for 20 minutes.

In technique II, a double digestion was performed using *Pst 1* followed by a secondary restriction enzyme, *Hae III,* cat. # R0108S (New England BioLabs, Ipswich), using the digestion protocol for sample C above. This enzyme is a four base cutter hence will cut approximately every 256 bases with a 5'-GG^CC-3' recognition site. The double digestion was performed to digest the gDNA sample to an approximate length of 300 bp without destroying any of the *Pst 1* recognition sites while reducing the fragment length to the required without fragmentation. Digested samples A, B and C were then purified with Qiagen MinElute PCR purification kit columns, cat. # 28004, (Qaigen,Valencia) and eluted in 15 µL, 35 µL and 35 µL of nuclease free water respectfully. Samples B and C were then stored at a temperature range of 2^{0}C to 8^{0}C for later processing.

Fragmentation

The 15 *Pst 1* – digested genomic 'sample A' was then further fragmented using NEBNext dsDNA fragmentase, cat. # M0348S, (New England BioLabs, Ipswich). This enzyme cuts DNA randomly and final fragment is dependent on the initial concentration of the DNA and the incubation time for the reaction. The reaction mixture contained 2 µL of Fragmentase enzyme, 2 µL of 10X reaction buffer, 0.2 µL of BSA and the final volume was raised to 20 µL with nuclease free water. This reaction was incubated at 37^0C for 20 minutes in order to gain an average fragment size of 300 bp which is very close to the optimal 260 bp suggested by Ion Torrent for sequencing 200 bp reads. The reaction was terminated by adding 5 µl of 0.5M EDTA. The sample was then purified with Qiagen MinElute PCR purification kit columns, cat. # 28004, (Qaigen, Venlo) while paying close attention to the pH of the fragmented DNA sample. This was usually achieved by adding 10 µL of 3M sodium acetate to the sample when the pH was observed to be too high in order to facilitate adsorption of DNA to the silica gels in the columns. A final clean sample volume of 35 µL of was then stored on ice. This fragmentation step was performed on only sample A.

Adaptor Design

Two types of adaptors are used in developing DNA libraries for sequencing on the Ion Torrent PGM, namely, "A-adaptor" and the "P-adaptor." Both adaptors contain known PCR priming sites. The "A- adaptor" which has the recognition sequence (5'-TCAG-3') for the Ion Torrent PGM was modified for directed ligation to the *Pst 1* digested genomic DNA. This was achieved by designing and ordering separate top and bottom oligos of the "A-adaptor" fragment with the top oligos having extra nucleotides

that are complementary to the *Pst 1* restriction overhang on the digested genomic fragments. A 100uM "A-adaptor" stock solution consisting of equimolar top and bottom strand, 5'- CCATCTCATCCCTGCGTGTCTCCGACTCAGTGC*A-3'and 5'phos CTGAGTCGGCGACACACGGGGATGAGATGG-3 respectively, were denatured at 98^{o}C and the temperature was decreased by 1^{o}C per minute until 25^{o}C in order to promote effective and accurate complementary annealing of the oligos (Amores et al., 2011). The P-adaptor sequence (5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTG AT-3', 5'-ATCACCGACTGCCCATAGAGAGGAAAGCGGAGGCGTAGTGG*T*T-3) has a complementary strand of DNA on the Ion Sphere Particles (ISPs) which is used to amplify the library through emulsion polymerase chain reaction, emPCR, prior to loading onto the chip for sequencing. The P - adaptor was obtained from the Ion Express 1-16 barcode kit, cat. # 4471250, (Life Technologies, Carlsbad). The "A-adaptor" and "P-adaptor" fragments were compared to the standard A and P adaptor mix from the NEBNext® Fast DNA Library Prep Set for Ion Torrent (NEB # E6270S/L) (New England BioLabs, Ipswich) using the Agilent 2200 TapeStation System, cat. # G2964AA, and the D1K Screen Tape, part # 5067-5361 (Agilent Technologies, Santa Clara).

A-Adaptor Ligation

A total amount of 5 µL of the modified ion torrent "A-adaptor" (200nM) was then ligated unto the *Pst 1* digested and fragmented sample A and double digested sample B. A 50 µL reaction contained 3 µL of T4 ligase, 5 µL of T4 ligase buffer and 0.5 µL of BST for each sample. Finally, sample volumes were adjusted with nuclease free water and 1.5 µL of ATP was added to begin the reaction. The mixtures were incubated at a

25°C for one hour. The ligase enzyme was deactivated by incubating at 65^{0}C for 30 minutes. The samples were then purified using Qiagen MinElute PCR purification kit columns, cat. # 28004, (Qaigen,Venlo) and eluted in 35 µL of nuclease free water.

In order to verify that the A-adaptor ligated to the *Pst I* restriction digested genomic DNA overhangs, a ligation trial of the modified adaptor was performed. A genomic DNA sample was digested with *Pst I* then ligated to the modified A-adaptor as described above. The sample DNA was then amplified for 30 cycles (95°C for 1 minute, 64^{0}C for 30 seconds, 72^{0}C for 4 minutes). A 4-minute extention time was set due to the length of the *Pst I* digested gDNA fragments, approximately 4kb. In order to effectively amplify the *Pst I* digested and ligated samples, the A-primer was used as both the forward and reverse primers (5'- CCATCTCATCCCTGCGTGTCTCCGACTCAGTGCA- 3').

End Repair

In order to facilitate the ligation of the "P-adaptor" an end repair reaction was then performed with the End-It™ DNA End-Repair Kit, cat. # ER0720 (Epicenter, Madison, WI-USA) on all three samples since the "P-adaptor" ligates onto only blunted ends. The 34 µL of the eluted "A-adaptor" samples (A and B) and digested sample C were used for end repair. Reactions contained 5 µL each of buffer, dNTPs mix, ATP and 1 µL of enzyme from End-It™ DNA End-Repair Kit (Epicenter, Madison, WI-USA). The final volume was adjusted to 50 µL and incubated in a thermal cycler for 45 minutes at 25°C. The enzyme was deactivated by holding the samples at 70°C for 10 minutes.

P –Adaptor Ligation

A similar procedure as that for the "A-adaptor" ligation was used for "P-adaptor" ligation. Ten µL ligase buffer and 10 µL P-adaptor were added to the 50 µL end repaired

samples (A and B). A final reaction volume of 100 was achieved by adding 6 µL of T4 ligase enzyme, 1 µL of BST, nuclease free water and incubated for 15 minutes at 25^{o}C. The enzyme was then deactivated by raising the temperature to 65^{o}C for 5 minutes and centrifuged before storing on ice.

For sample C, the standard adaptor mix from NEBNext® Fast DNA Library Prep Set for Ion Torrent (NEB # E6270S/L) (New England BioLabs, Ipswich) was used to randomly incorporate the A and P adaptors unto the blunt ends of the *Pst I* digested sample C. This was because there was no fragmentation and "A-adaptor ligation prior to end-repair A 100 µL reaction mixture comprising 50 µL of blunt-end sample C, 10 µL of T4 ligase buffer, 20 µL of adaptor mix, 6 µL of enzyme, 1 µL BST and nuclease free water was incubated for 15 minutes at 25^{o}C with a deactivation step of temperature 65^{o}C for 5 minutes, followed by centrifugation and storage on ice.

Size Selection

Size selection was performed to select for the total fragment length to be amplified to the recommended size range, i.e. adaptors and target fragments. Agencourt AMPure XP 5 mL Kit, cat. # A63880 (Beckman Coulter, Brea CA-USA), was used for size selection of ligated sample fragments ranging between 310 and 370bp from all three samples. The beads were first vortexed for 30 seconds before use. Following kit protocols, fragments larger than 370 bp were removed by adding a volume of Ampure beads equivalent to 70% of the sample. The mixture was mixed thoroughly, incubated for 5 minutes at room temperature and then set on a magnetic rack for approximately 5 minutes. The supernatant was then collected without disturbing the beads at the bottom. Subsequently, 15 µL of beads (15% vol/vol of initial DNA sample) were then added to

the supernatant to recover fragments of 310-370 bp. The mixture was placed on the magnetic rack and the clear supernatant was collected and discarded. Beads bound to desired size range fragments were washed twice with freshly prepared 80% ethanol while the sample bound beads were still bound on the magnetic rack. After removing and discarding the alcohol, the sample was allowed to air dry. The sample bound beads were subsequently eluted in 45 µL nuclease free water and allowed to incubate at room temperature for approximately 5 minutes while off the magnetic rack. The elution was then put back on the magnetic rack for about 3 minutes to separate the beads from the solvent. A total of 40 µL of solvent was carefully removed so as not to disturb the beads since the beads had no DNA bound to them.

Amplification

The size selected fragments were then PCR amplified to increase the number of acquired sequence library fragments for downstream sequence reactions. PCR was stopped while still in the linear phase. Reagents and primers from the NEBNext® Fast DNA Library Prep Set for Ion Torrent™, cat. # E6270S (New England BioLabs, Ipswich) were used amplification. The kit protocol was used for initialization, denaturation, annealing and extension but the number of cycles was increased to 8 to make sure the RADTag libraries amplified. After the amplification reaction, the sample was purified with Ampure beads using a bead to DNA ratio of 1:1. Amplified samples were then eluted in 35 µL of 100mM TE buffer for storage and subsequent sequencing reactions.

Template Preparation

Template preparation aids in the ligation of the library to ION Sphere Particles (ISPs), amplification of ISP-Template conjugate and then denaturing the DNA to obtain single strand templates. This step is very important as it determines quality of sequences and also makes the fragments recognizable to the ION Torrent PGM.

Emulsion PCR

Prior to the emulsion PCR (emPCR), a partial genomic DNA library of *E. coli* from the Ion Control Material 200kit, cat. # 4471249 (Life Technologies, Carlsbad) with known concentration and fragment size was diluted to approximately 26 pmol/L. This control was used following Kit protocols in determining the average concentration of samples in ng/L needed for an effective emulsion PCR. The samples (A, B, and C) were then diluted with nuclease free water in order to gain less than 26 pmol/L of DNA. An amount of 20 µL of the diluted sample was then used in each sample's emPCR procedure. A 1000 µL amplification solution made up of Ion OneTouch™ 2X Reagent Mix, Ion OneTouch™ Enzyme Mix and Ion OneTouch™ 200 Ion Sphere™ Particles from the Ion OneTouch™ 200 Template Kit v2 DL, cat. # 4480285, (Life Technologies, Carlsbad), together with the freshly diluted sample prepared and loaded into the Ion OneTouch DL instrument (Life Technologies, Carlsbad). Prior to loading the amplification solution, 1500 µL of Ion OneTouch reagent oil was added. The protocol for amplification solution prep for the Ion OneTouch™ 200 Template Kit v2 DL, cat. # 4480285, revision 6 was followed. The Ion Sphere Particles (ISP) ligated and amplified fragments (ISP-Template Conjugates)were then assessed with an Ion Sphere Quality Control Kit (Ref # 4468657) on the Qubit® 2.0 Fluorometer (Sku # Q32866) (Life

Technologies, Carlsbad), to determine the ISP-Template Conjugates prior to enrichment on the Ion OneTouch™ ES (Life Technologies, Carlsbad). The ISP-Template Conjugates fragments were then enriched on the Ion OneTouch ES. This was achieved by denaturing the ISP-Template Conjugates using a "Meltoff" solution made up of 1M NaOH, 10% Tween and nuclease free water as specified in the Ion OneTouch™ 200 Template Kit v2 DL user guide and Dynabeads MyOne Streptavidin C1 Beads, cat. # 650.01, (Life Technologies, Carlsbad).

Sequencing

The Ion PGM™ Sequencing 200 Kit v2, cat. # 4482006, (Life Technologies, Carlsbad) was used to sequence prepare the ISP-Template conjugates as described in the Ion PGM™ Sequencing 200 Kit v2 User Guide, cat. # 4482006, revision 3. Sequencing was performed on the Ion Torrent Personal Genome Machine (PGM) (Life Technologies, Carlsbad) using a 316 chip.

Analysis

Genomic sequence data from the Ion Torrent PGM was recovered on the Ion Torrent browser. A comparison was made between our sample data and a published mtDNA of the same species (Lee et al*.,* 2001) to verify the amount of mitochondria DNA that were still present in our Genomic samples. Submission of genome sequence data to a public database is pending. The Stacks software (Catchen et al., 2013) was used to clean the sequence data by trimming all the sequence reads to 100 bases and by rejecting reads of low quality since the software works optimally with sequence fragments of equal lengths. It was also used to identify and analyze specific loci markers in the sequences such as SNPs. Sequenced samples A, B and C were analyzed individually. Pooled

genomic samples A, B and C were also analyzed for comparison with individual samples to check for representative sampling of sites across the genome of all three library prep methods. Four major pipelines in the stacks software were employed in the RADTag genome sequence data, namely:

- Process_radtags – this cleans and trims the input sequences
- Ustacks – which groups similar sequences into groups called ‘stacks’
- Cstacks – builds catalogues of the stacks by comparing one stack to another
- Sstacks – analyses samples of interests using built catalogues as a reference
- Populations – this command treats the samples as members of a population while looking for similarities and differences as well as conducting population statistic evaluations.

Chapter III

RESULTS

Genomic DNA Extraction

Extracted gDNA was relatively pure and free from contamination and RNA with a 260:280 spectrometer ratio of 1.9. An average length was approximately 18 kb with a concentration of greater than 199 ng/µL was observed for Hon 9 genomic DNA isolates (Figure 5).

Digestion

Analysis of digested samples on the bioanalyzer demonstrated *Pst 1* did digest samples effectively from lengths greater than 18kbp to an average of 5307 bp (Sample A) and 1389 bp (Sample C) with a concentration of 6.22 ng/l and 9.57 ng/µL respectively (Figure 6). Since *Pst I* is theoretically expected to cut just about every 4,096 base pairs the average observed fragment size of sample A is close to the predicted. The double digestion of Sample B produced an average fragment length of 658 base pairs (Figure 6).

Fragmentation

NEBNext dsDNA fragmentase, cat. # M0348S, (New England BioLabs, Ipswich) is expected to yield a fragment length ranging from 200-1000 bp with a reaction time of 15–25 minutes. A fragmentation reaction time of 20 minutes resulted in a partially fragmented sample with an average fragment length of 306 bp (Figure 7).

Ligation

A comparison of our modified A-adaptor and Ion Express P-adaptor with the standard adaptor mixture from 'NEBNext Fast DNA Library Prep Set for Ion Torrent' showed that our Adaptors were similar in fragment length and quality (Figure 8). A difference was observed in the molar concentrations which were adjusted empirically before using our adaptors. We adjusted our A-adaptor from 200 nM in all subsequent reactions based on this result. The modified adaptor ligation test resulted in amplified fragment of approximately 3000 bp in length while the expected fragment length was 4096 base pairs (Figure 9).

Amplification

Linear amplification was a success, yielding more than enough product concentration for adequate dilutions for the next reactions. This also verified that the size selection protocol worked (Figure 10). Samples varied in both concentrations and fragment lengths with Samples A and B having concentrations less than 2.5 ng/µL with an average fragment length of 281 bp. A high DNA concentration of 25.7 ng/µL was observed for Sample C with an average fragment length of 326 bp (Table 2).

Emulsion Amplification

Emulsion amplified and ISP-Template conjugate sample, were analyzed with the Ion Sphere Quality Control Kit, Ref # 4468657, ((Life Technologies, Waltham) on the Qubit® 2.0 Fluorometer, Sku # Q32866 (Life Technologies, Waltham)and a reading of 24, 23.9 and 18.5% of ISP-Template conjugate were recorded for Samples A, B and C respectively (Table 3). These results fall within the recommended optimal range (10-30%) by Ion Torrent in order to minimize ISP-Template conjugate that may result from

the polyclonal fragments. Enrichment of ISP-Template conjugate were observed to be efficient given that almost a 100% of the ISP-Template conjugate loaded for sequencing were enriched (Table 4).

Sequencing

Initial analysis of sequences on the Ion Torrent browser showed high quality sequence reads with a total of 552 M, 581 M and 738 M sequenced bases for Samples A, B and C, respectively. There was a positive correlation between loading efficiency and sequencing output provided template preparation was performed adequately (Figure 5). The average sequenced genomic fragments length was 175 bp, 162 bp and 191 bp for Samples A, B and C, respectively with most of the sequenced fragments length being greater than 180 bp. Aligning samples sequences with a reference mitochondria sequence showed less than 10% of the genomic DNA to be mtDNA across each of the genomic samples (Table 5) with about 100X coverage for Samples A and B, and 245X coverage for Sample C. All sequence fragments began with the *PST I* recognition site (5'-TCGA-3') with the exception of Sample C since Sample C is consists of both RADTag and regular sequences.

Analysis

Analysis of the sequences from samples with the RADTag software (Stacks) identified lots of SNP markers with the exception sample C. Individual analysis of the samples as individual populations showed thousands of variant sites with the exception of Sample C showing very few variant sites (Table 6). In Samples A and B, about 9000 and 7000 SNPs were identified respectively, but only 13 SNPs in Sample C. Heterozygosity was observed to be very low (< 0.01) in all three samples with homozygosity score of

0.999. Comparisons of pooled sequence reads with individual samples produced similar results as observed in the individual sample analyses (Table 7).

Chapter IV

DISCUSSION

Separation of Mitochondrial and the Nuclear Genomic DNA

Mitochondrial DNA is very easy to work with due to its stability and size. Certain hindrances such as insertion activity of mitochondrial DNA sequences into the nuclear DNA (Hazkani-Covo, et al*.,* 2010) have been observed more frequently over recent years. This phenomenon in addition to the occurrence of mutations (insertion and deletions) make commonly used genotypic analytical methods such as AFLPs, RFLP and microsatellites flawed since conclusions from these techniques are drawn based on fragment length but not the actual sequences.

Furthermore, the insertion of mitochondria DNA into gDNA also influences the output of genomic DNA sequencing. About 40% of presumed genomic sequences are reported to actually be mitochondrial DNA in origin (Ring, personal communication). The cause of this high mtDNA percentage input may be attributed to the high copy number of mitochondrial DNA sequences relative to the genomic sequences along with the mitochondrial insertions in the nuclear genomic DNA. Research has shown that the actual amount of mitochondrial DNA insertion into gDNA is unpredictable. As high as 25% heterozygosity in *Cyt b* of primates (Collura & Stewart, 1995) and as low as 1% heterozygosity in *COI* and *COII* of aphids (Sunnucks & Hales, 1996) has been reported when mtDNA are compared with their gDNA insertions.

These unpredictable occurrences among other factors required the development of a method for the separation of the nuclei from mitochondria prior to DNA extraction in order to reduce the mtDNA contamination of our final genomic sequence data. Analysis of all three samples using the published mitochondria sequence as a reference showed that less than 10% of our fixed sequences were mitochondrial. Also, purely RADTag samples (Samples A and B), had as low as approximately 5% of the entire sequence as being mitochondria DNA. This is half of what is observed in Sample C which consisted of RADTags and regular genomic sequences. It can therefore be concluded that the separation of mtDNA and gDNA influenced the high specificity (90%) of our genomic sequence output a (Table 4).

Library

Cresko labs, the developers of RADTag sequencing employed the use of the digestion enzyme *Sbf I,* which is an eight base cutter and cuts approximately every 65kb. This enzyme was used because it was known to have about 30,000 recognition sites in the typical teleost fish genome (Amores et al*.,* 2011). It was decided to use *PST I* instead of *SBFI* because of its increased number of recognition sites (approximately every 4 kb) which will provide a means to cover more of the genome under study.

The fragmentation reaction is performed to cut the gDNA fragment into the appropriate lengths suitable for the sequencing platform. When fragmentase enzyme was used (Sample A), it cut randomly along the genome without any regards to specific DNA sites. In the RADTag protocol developed for Solexa (Illumina, San Diego) (Amores et al*.,* 2011), fragmentation was performed post ligation of the adaptor. This was modified in our protocol in order to reduce the damage of important fragments and adaptors, by

fragmenting prior to the ligation of the adaptors. This helped to preserve these essential adaptors which finally greatly improved the overall output of sequencing data. The idea of fragmentation damaging some of the required DNA brought about the initiative of exempting it all together in Samples B and C. A double digestion was performed on sample B with *PST I* as well as *Hae III*, cat. # R0108S (New England BioLabs, Ipswich) as the enzymes. *Hae III*, cat. # R0108S (New England BioLabs, Ipswich) digestion was opted for as a means of fragmenting the digested samples into the smaller sizes for library preparation and sequencing. Since *Hae III*, cat. # R0108S (New England BioLabs, Ipswich) is site specific, no restriction overhangs were destroyed and the *Hae III*, cat. # R0108S (New England BioLabs, Ipswich) ends can also be used for further analysis. In view of the fact that there was no fragmentation reaction for Sample C, the incubation period of the digestion reaction was increased to 2 hours to provide more time for the restriction enzyme *(PST I)* to locate and cut more sites. This resulted in smaller size fragments (Figure 2) when compared with fragments from Sample A (1 hour incubation time). Therefore, Samples B and C were size selected for the required length without the fragmentation reaction step.

End repair enzymes create blunted end DNA fragments by degrading overhangs or by filling up the gaps. This did not have any adverse effect on Samples A and B because the over hangs of the *Pst 1* digests were protected by the ligation of the modified A-adaptor. In Sample C, since there was no ligation of the A-adaptor prior to the end repair reaction, some of the over hangs of the Pst 1 digest were degraded by the end-it enzyme. In Sample C, the standard A & P adaptor mix from NEBNext® Fast DNA Library Prep Set for Ion Torrent, NEB # E6270S/L, (New England BioLabs, Ipswich)

was used in the ligation step after the end repair step. This adaptor mix ligates randomly onto blunted end hence it ligated to all bunted fragments, i.e. fragments with the *Pst 1* over hang nucleotides as well as any random fragment that has been blunted. This resulted in the final output of the library preparation for Sample C not being solely RADTag sequences but also regular genomic sequences. In addition, size selecting for fragment length ranging from 310-370 bases while the average *PstI* digested fragment is 1kb for Sample C, implied that the RADTag output was further limited and much of the genome was not be covered in this sample (Table 6).

Emulsion PCR

Emulsion PCR is the process that aids in binding the ISPs onto the prepared library and then amplifies the fragments. Here, the concentration of the input linear amplified gDNA library is very critical so as not to over amplify the fragments. Ideally, each ISP should have not more than one fragment attached to it prior to amplification. This will avoid the formation of polyclonal which affect the sequencing quality and output. The Ion OneTouch™ 200 Template Kit v2 DL user guide demands an input DNA of concentration 26 pmol/L in order to obtain an ISP-Template conjugate percentage of 10-30% which is the optimum range to avoid polyclonal. The research showed that aiming for this 26 pmol/L usually produced fragment bound ISP percentage over the optimum. This was attributed to the fact that minute pipetting errors can cause a great difference in the final concentration of the diluted amplified genomic DNA library in such minuscule concentrations as are needed for the emPCR. Hence a protocol was used that aimed for lower concentrations of 15, 15 and 13pmol/µL which yielded 24%, 24% and 18% ISP-Template conjugates for Samples A, B and C respectively (Table 3). This

proved successful in reducing the amount of polyclonal sequences in the final data (Table 4). It was also observed that this low molar concentration coincided with concentrations less than 3.5 pg/μL (Table 3). This implies that in certain cases, one might not have to perform the linear amplification process provided ligation and purification were done adequately.

Analyses of Sequences

Unlike sequencing with Solexa (Illumina, San Diego), sequenced reads from the Ion Torrent PGM (Life Technologies, Carlsbad) are not entirely of the same length (Table 5) hence, trimming was required to enable analysis to be done using the Stacks software which was designed to analyze RADTag sequences by aligning specific end sites. Samples A and B had mtDNA sequences of an average length less than 50 bases while Sample C had approximately 150 bp (Figure 7). Since trimming the sequences not only reduces longer fragment lengths to the arbitrarily chosen length, but also rejects all fragments lengths less than the chosen. An average fragment length of a 100 bases was selected for all three samples as this aided in reducing the influence of any mtDNA sequences (average length 50 for Samples A and B) on the analysis of RADTag sequences. Even though some mitochondria DNA may still be present in Sample C, it was rendered insignificant in the analysis since it would not align to any other sequence present.

Heterozygosity

Occasional out-crossing has been suggested to be the leading cause of heterozygosity in this self-fertilizing cryptic species, *K. marmoratus*. By analyzing few genetic markers through AFLP and microsatellites, conclusions on the heterozygous

nature of these fishes were drawn (Nakamura et al., 2008; Mackiewicz et al., 2006). Nevertheless, inbred organisms are expected to lose 50% of their genetic variation at every generation (Tatarenkov et al*.,* 2010), signifying that in the event of no out crossing, this self-fertilizing species will eventually return to its original largely homozygous nature within ten generations. However, several researchers have questioned the level of homozygosity existing among coding loci and how a supposedly highly homozygous organism has survived over the years. Our data depicts low levels of heterozygosity (0.0003) in all modes of comparison, that is, individual sample sequence comparison as well as pooled samples sequence comparison except for Sample C, which registered lower heterozygosity value (0.00014). For the Sample C sequences, the lower heterozygosity was recorded due to the fact that most of the sequence data were regular partial genomic sequences with few sites (89,829) being RADtag sequences therefore resulting in as low as 13 polymorphic sites. With respect to the unheard of low levels of heterozygosity in all analyzed scenarios, it might be suggested that the *K. marmoratus* has adapted a mechanism that reduces random mutations tremendously. This result is feasible since the benefits of inbreeding (self-fertilization) are speculated to out-way that of out crossing. It has been suggested that self-fertilization reduces the cost of mate search (Tatarenkov et al*.,* 2009) and also aids in fixing beneficial traits for survival (Allard, 1975).

The use of RADTag sequences and Stacks analyses aided in locating thousands of SNP markers which can be used for other in-depth analyses (Appendix C). Here we identify 8,000 such loci. This provides the opportunity to make genomic comparisons on a vastly broader scale. Results from the RADtag sequences confirm that *K. marmoratus*

exhibited high numbers of variant sites, but these sites were still low in comparison to the total fixed sites analyzed (Tables 6 and 7). Therefore, though heterozygosity can be observed in *K.marmoratus*, analyzing a broad range of markers produces a better picture of the species' genotypic characteristics. Given that SNPs are very informative on relatedness of organisms on the individual and the population level, our battery of 8,000 such markers should prove invaluable to taxonomic and population studies in the coming future.

Conclusion

Though fragment length analysis is relatively less expensive and time conducive, its major flaw is that it provides a very limited survey of genomic information that can drastically limit the output and affect conclusions made from these results. Though the identification and building of markers on the Ion Torrent PGM through RADTag sequencing might be more expensive than common fragment based methodologies, it has many benefits in that identified RADTag sequences can be used for many other significant analyses. Not only is RADTag sequencing cost effective on the Ion Torrent PGM, this system, using our methods, yields vastly more data resulting from more efficient sequencing of more fragments of longer sequence length than existing methods on Next Generation Sequencing systems. The Ion OneTouch System, Ion OneTouch ES and the PGM are very user friendly while still producing high outputs with high accuracy. Therefore, analysis is made easier while exposing thousands of variant sites.

BIBLIOGRAPHY

Allard, R. W. (1975). The mating system and microevolution. *Genetics*, 79, 115-126.

Amores, A., Catchen, J., Ferrara, A., Fontenot, Q. & Postlethwait, J. H. (2011). Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188(4), 799-808.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., Ingram, K & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148-155.

Catchen J., Hohenlohe P., Bassham S., Amores A. & Cresko, W. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124-3140

Chandel Lab. Mitochondria Isolation Protocol (Internet). Available from http://www.chandellab.com/Protocols/.

Collura, R. V. & Stewart, C. B. (1995). Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature*, 378, 485-489

Costa, W. J. E. M. (2004). Kryptolebias, a substitute name for Cryptolebias Costa, 2004 and Kryptolebiatinae, a substitute name for Cryptolebiatinae Costa, 2004 (Cyprinodontiformes: Rivulidae). Neotropical Ichthyology, 2(2), 107-108. Retrieved June 30, 2014, from http://www.scielo.br/scielo.php?script=sci arttext&pid=S1679-62252004000200009&lng=en&tlng=en. 10.1590/S1679-62252004000200009

Costa, W. J. E. M. (2004a). Relationships and redescription of *Fundulus brasiliensis* (Cyprinodontiformes: Rivulidae), with description of a new genus and notes on

the classification of the Aplocheiloidei. *Ichthyological Exploration of Freshwaters*, 15: 105-120.

de Knijff, P. (2014). How carrion and hooded crows defeat Linnaeus's curse. *Science* 344: 1345-1346.

Green, M. R. and Sambrok, J. (2010). Molecular Cloning; A Laboratory Manual. 4th edition. Cold Spring Habor Laboratory Press, Cold Spring Habor, NY.

Hazkani-Covo, E., Zeller, R. M. & Martin, W. (2010). Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genetics*, 6(2), e1000834.

Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H. & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14812-14817.

Hensel, R. (1868). Beitrage zur Kenntnis der Wirbelthiere Sudbrasiliens. *Archiv fur Naturgeschichte Wiegman*, 34, 323-375.

Hillis, D. M., Mable, B. K. & Moritz, C. (1996). Applications of molecular systematics: The state of the field anda look to the future. In: Hillis, D.M., Moritz, C. and Mable, B. K. Eds., *Molecular Systematics, Sinauer Associates, Massachusetts*, 515-543.

Mackiewicz, M., Tatarenkov, A., Taylor, D. S., Turner, B. J. & Avise, J. C. (2006). Extensive outcrossing and androdioecy in a vertebrate species that otherwise reproduces as a self-fertilizing hermaphrodite. *Proceedings of the National Academy of Sciences*, *103*(26), 9924-9928.

Nakamura, Y., Suga, K., Sakakura, Y., Sakamoto, T. & Hagiwara, A. (2008). Genetic and growth differences in the outcrossings between two clonal strains of the self-fertilizing mangrove killifish. *Canadian Journal of Zoology*, 86(9), 976-982.

Pfenninger, M. & Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology, 7*(1), 121.

Santos, F. P. S. (1997). The effect of body size on prey choice by *Rivulus luelingi* Seegers 1984 (Aplocheiloidei: Rivulidae). *Revista Brasileira de Biologia*, *57*(4), 551-562.

Sunnucks, P. & Hales, D. F. (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus Sitobion (Hemiptera: Aphididae). *Molecular Biology and Evolution*, 13(3), 510-524.

Tanya, C. & Kumar, R. (2010). Molecular markers and their applications in fisheries and aquaculture. *Advances in Bioscience and Biotechnology*, *Advance in Bioscience and Biotechnology* 1(4), 281-291.

Tatarenkov, A., Lima, S. M., Taylor, D. S. & Avise, J. C. (2009). Long-term retention of self-fertilization in a fish clade. *Proceedings of the National Academy of Sciences*, 106(34), 14456-14459.

Tatarenkov, A., Ring, B. C., Elder, J. F., Bechler, D. L. & Avise, J. C. (2010). Genetic composition of laboratory stocks of the self-fertilizing fish *Kryptolebias marmoratus*: a valuable resource for experimental research. *PLoS One*, *5*(9), e12863.

APPENDIX A

Figures 1 through 11

	Kryptolebias marmoratus	*Kryptolebias ocellatus*
Male		
Hermaphrodite		

Figure 1: Examples of two self-fertilizing and androdioecious vertebrates that can be regarded as cryptic species.
Image by F. Vermeulen

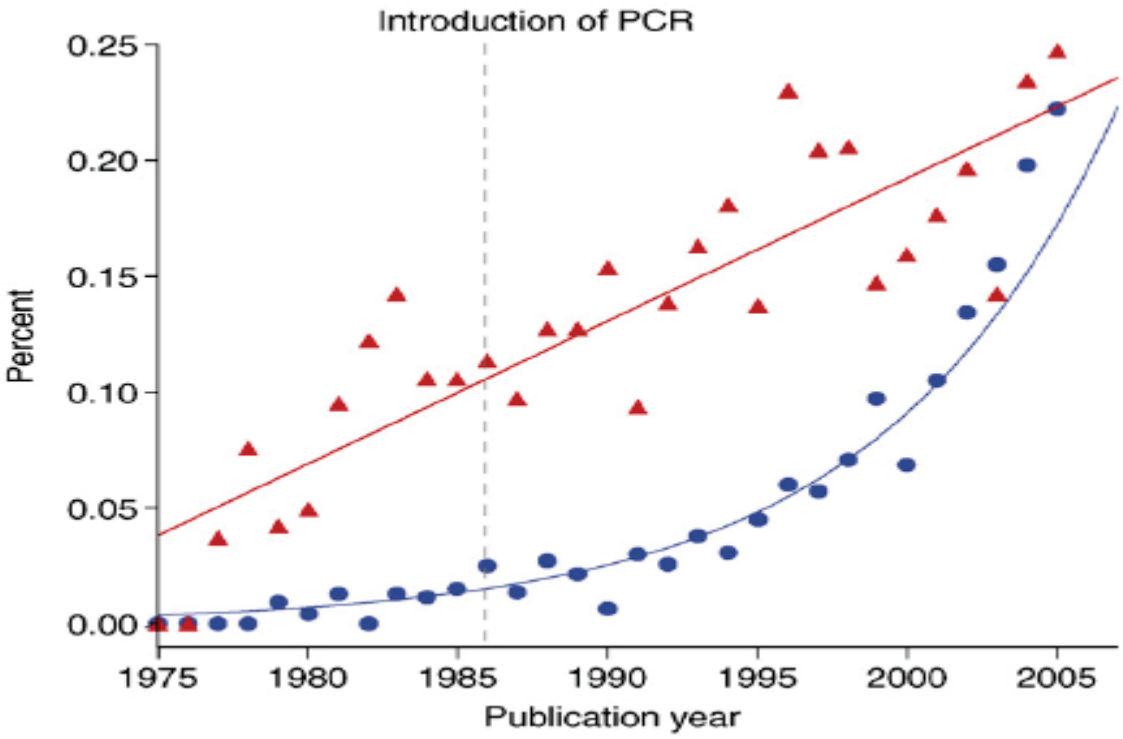


Figure 2: Increased recognition of Cryptic Species. The percent of peer-reviewed publications in Zoological Record Plus (CSA) that mention 'cryptic species' (circles) or 'sibling species'(triangles) in the title, abstract, or keywords has increased dramatically since the advent of PCR. Similar positive trends are observed in absolute number of publications per year, and in publications cited in other searchable databases of biological literature, including Science Citation Index (ISI) and Biosis Previews (Biological Abstracts) (OVID) (Bickford, 2007).

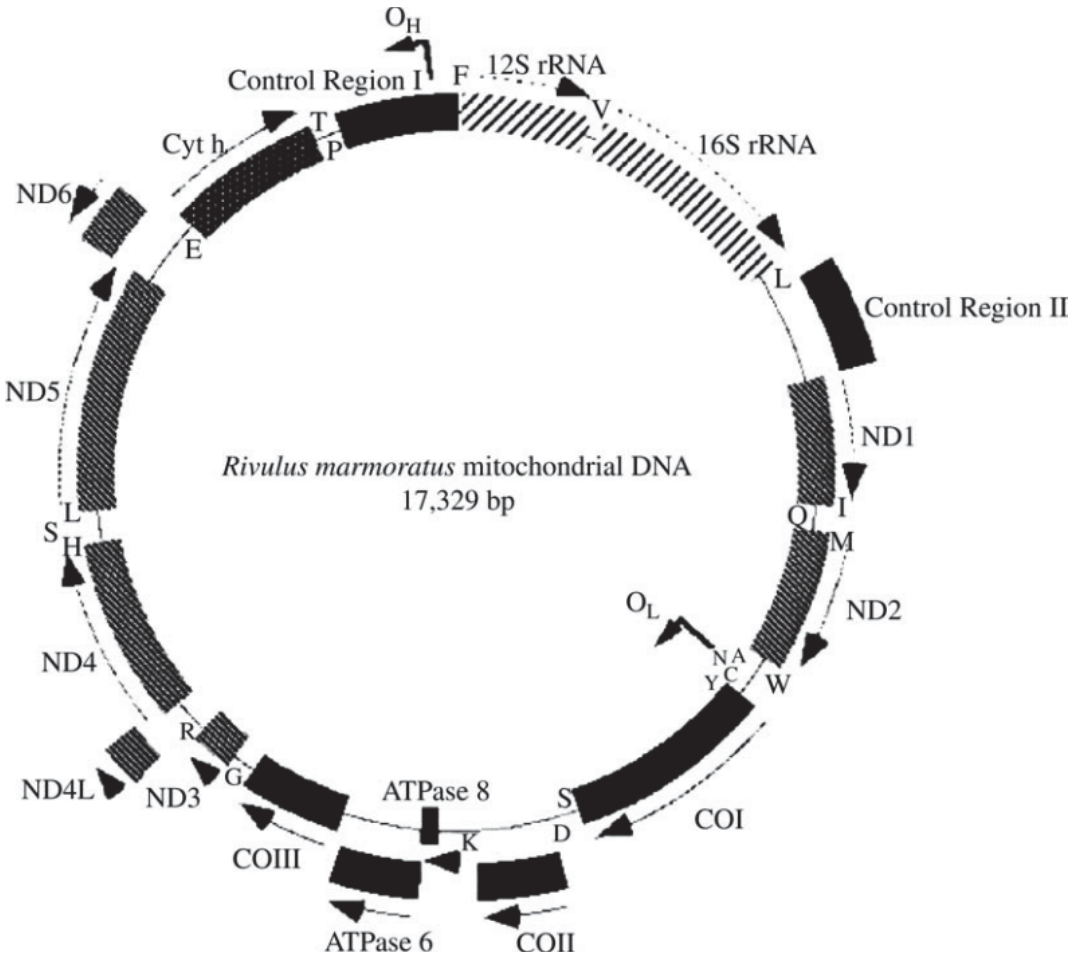


Figure 3: Mitochondrial DNA of *Kryptolebias marmoratus*.Complete L-strand nucleotide sequence (GenBank accession number: AF283503) and gene organization of the *Kryptolebias marmoratus* mitochondrial genome Lee et al*.,* 2001).

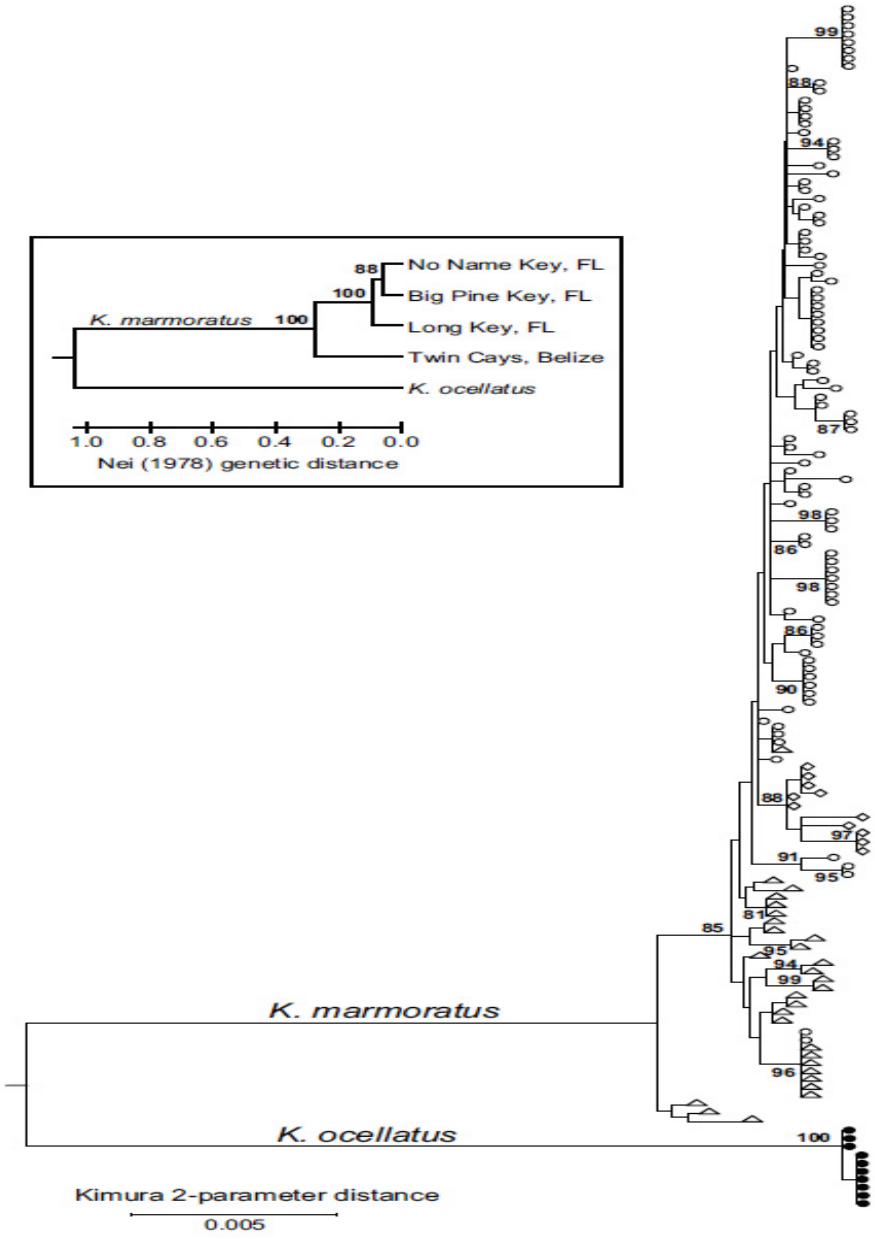


Figure 4: Genealogy of individuals based on Mitochondrial DNA. Genealogy for 136 individuals of *K. marmoratus* and 10 individuals *of K. ocellatus* based on 2,946-bp mtDNA sequences. Each circle, triangle, or rhombus represents an individual. In the *K. marmoratus* clade, triangles, open circles, and rhombi designate fish from Belize, various locations in Florida, and the Bahamas, respectively. Bootstrap values above 80% are shown. (Inset) Population phenogram for these species based on a cluster analysis of Nei's genetic distances from 31 microsatellite loci (Tatarenkov et al*.,* 2009).

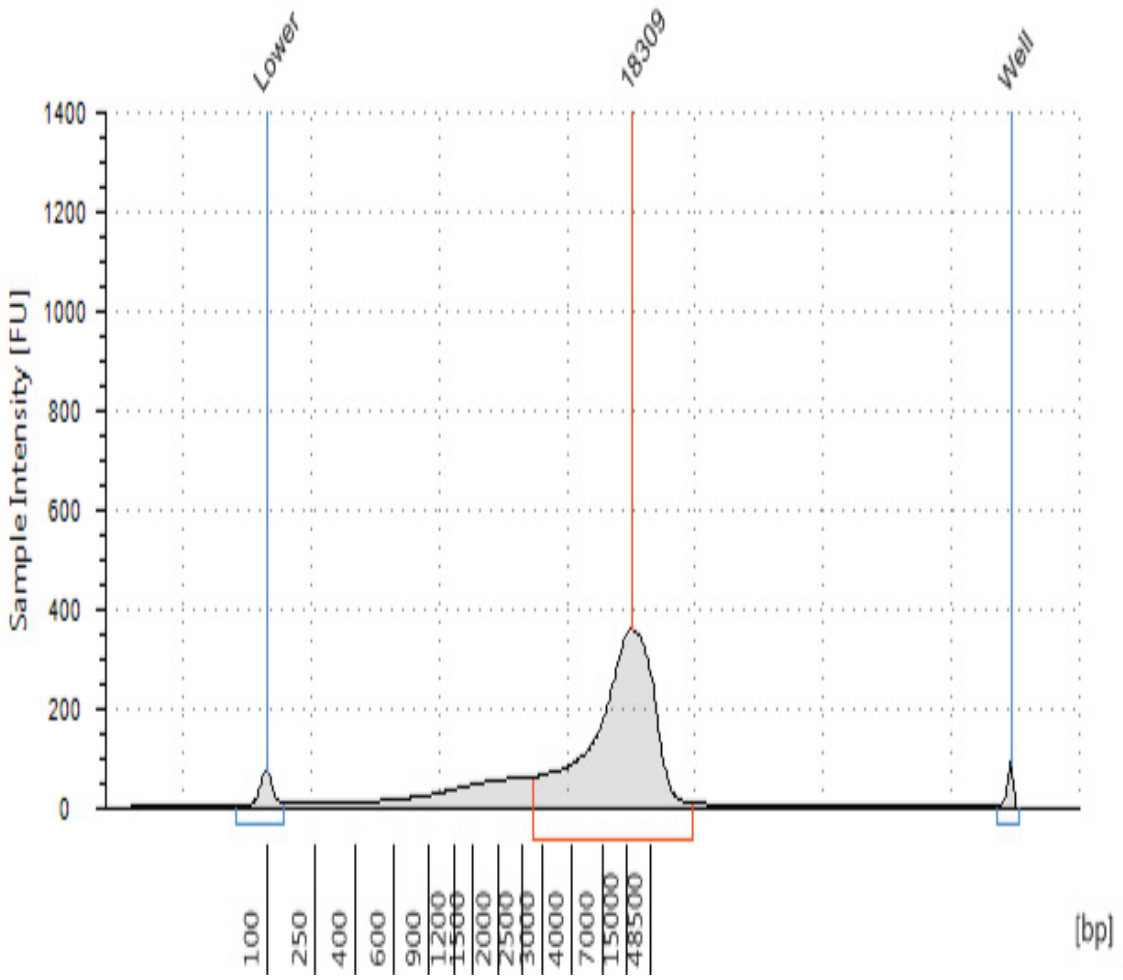


Figure 5: Analysis of extracted Genomic DNA. An electrophoregram showing the size distribution of extracted gDNA from Hon9 strain. The extracted genomic DNA had an average fragment length of 18309 with a DNA concentration of 199ng/µL.

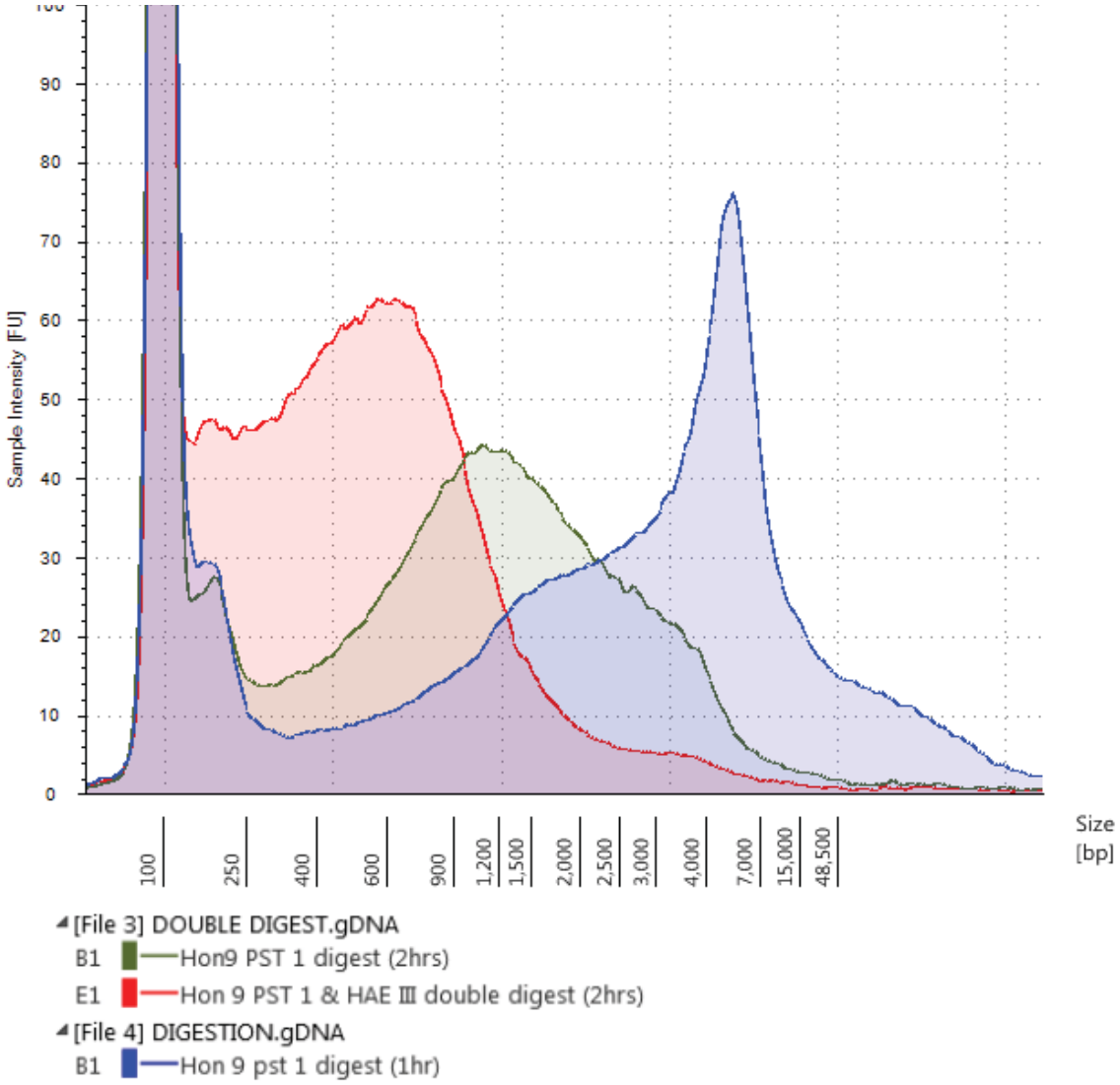


Figure 6: Analysis of Digested Genomic DNA Samples. Sample 'A' (blue) had an average of 5,307bp with a concentration of 6.22ng/µL after 1 hour of digestion with *Pst 1*. Sample 'B' (red), had an average fragmented length of 658bp with a concentration of 11.3ng/µL after double digesting for 2 hours. Sample 'C' (green) had an average fragmented length of 1389 with a DNA concentration of 9.57ng/µL after 1 hour of digestion with *Pst 1*.

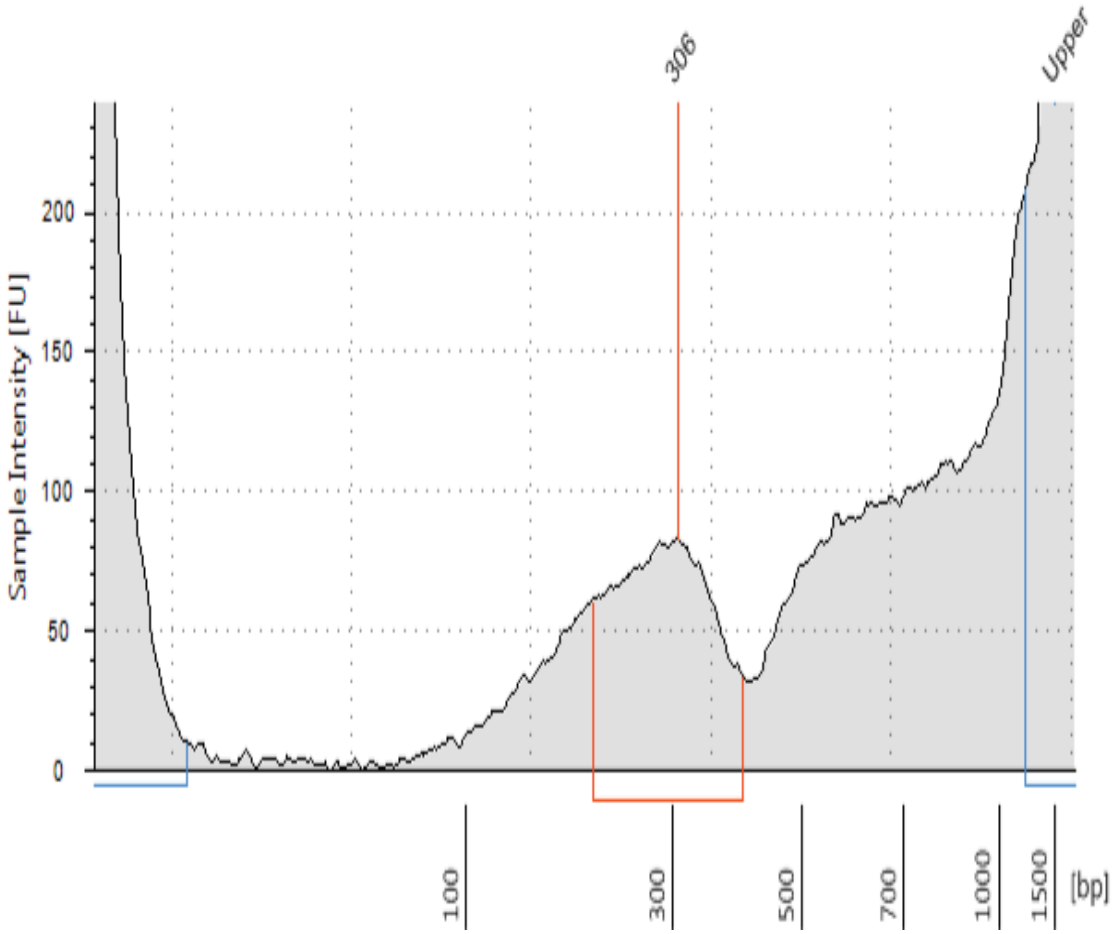


Figure 7: Analysis of fragmented Sample. An electrophoregram showing the size distribution of *Pst I* digested - Sample A after fragmentation with NEBNext dsFragmentase. Fragmented Sample 'A' had an average length of 306 with a DNA concentration of 1.07ng/µL

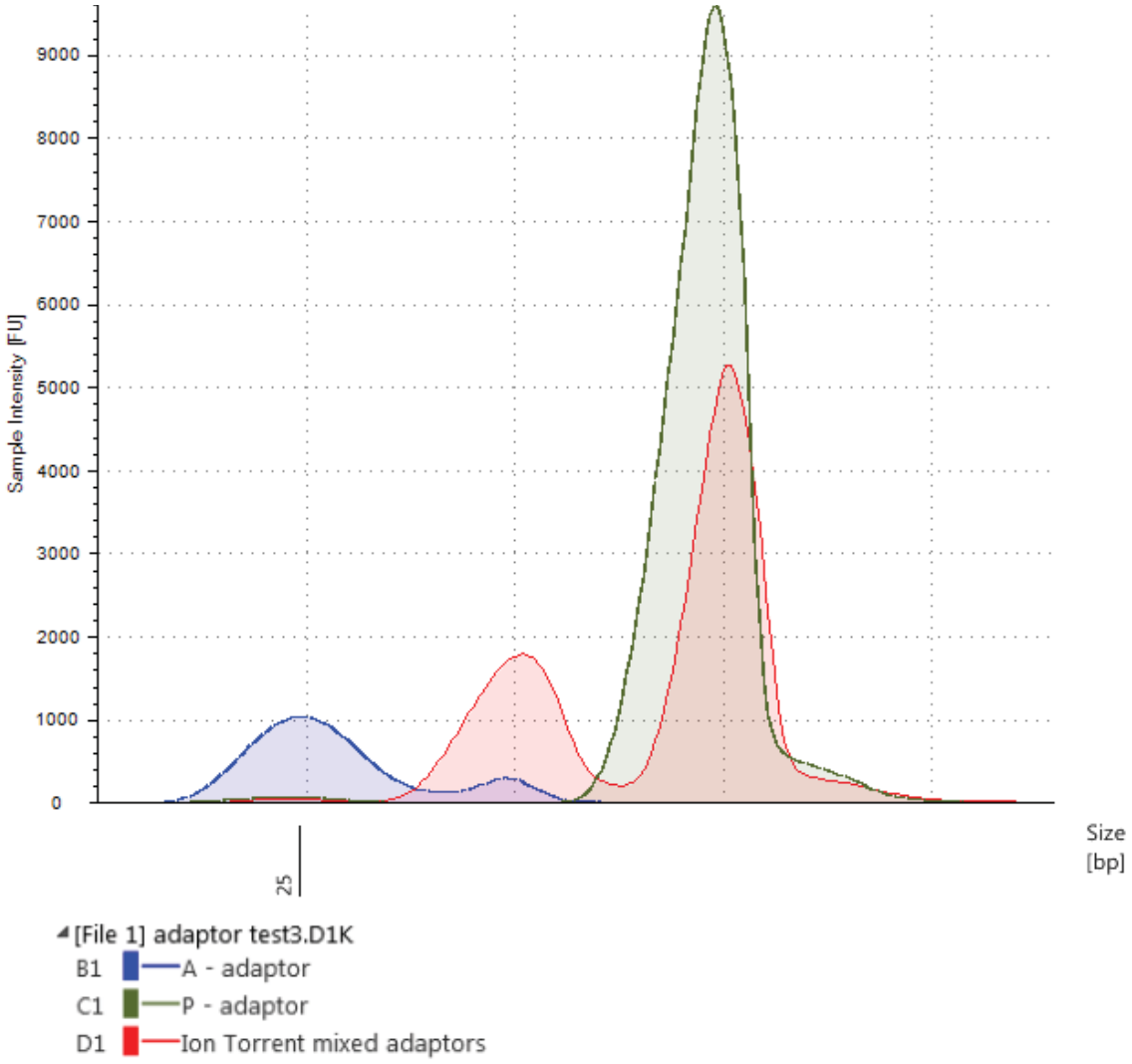


Figure 8: A Comparison of Adaptors. A comparison of the modified A-adaptor (blue) and the P-adaptor (green) from Ion torrent express barcode kit with the A&P adaptor mix (red) from the NEBNext® Fast DNA Library Prep Set for Ion Torrent showed similarities in fragment size. Differences were observed in the concentration of the adaptors.

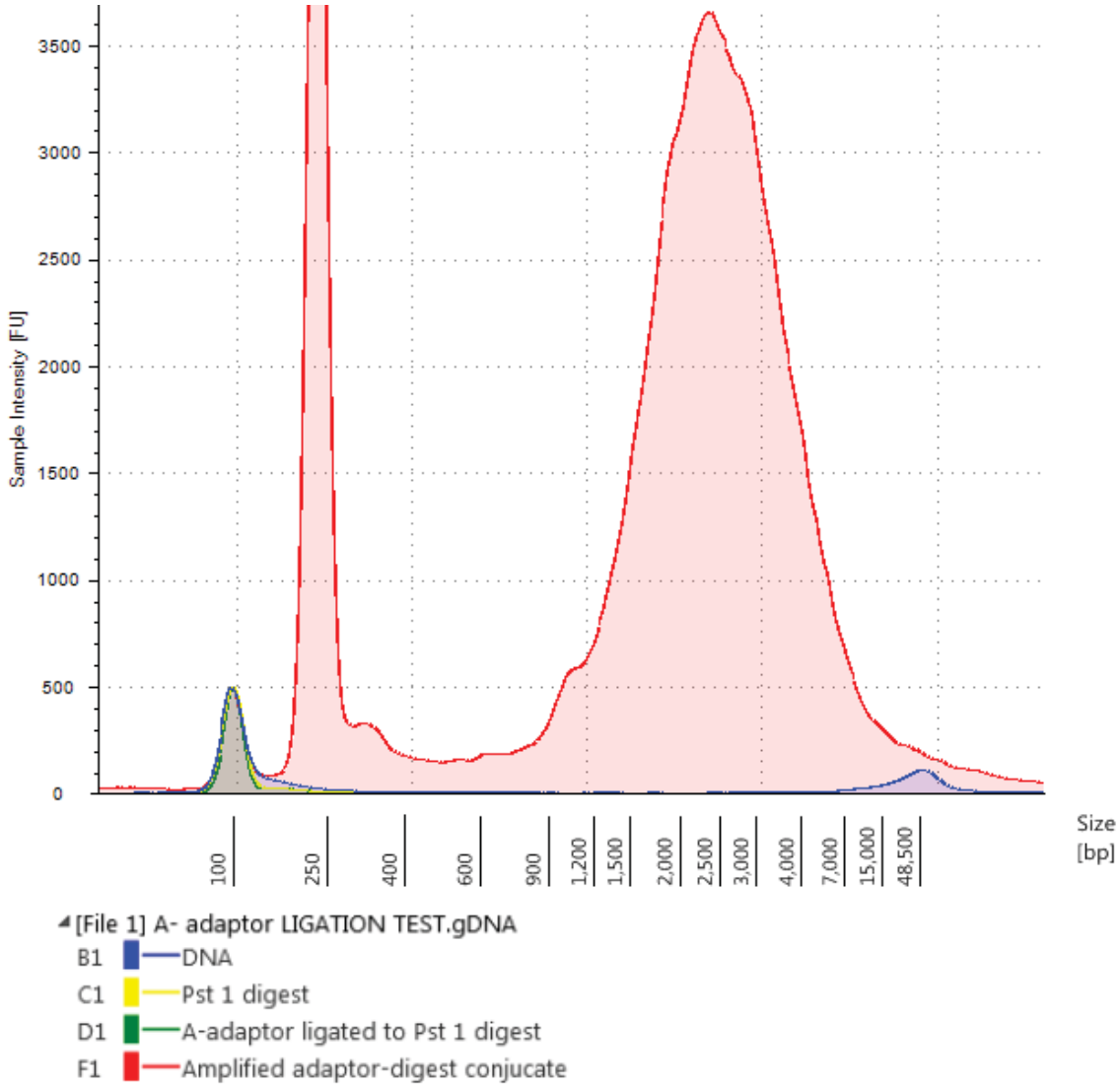


Figure 9: Modified A-Adaptor Ligation Test. An electrophoregram showing the various fragment size distributions of all processes involved in the A-adaptor ligation test. The average amplified A-adaptor ligated *Pst I* digested sample length was 2688bp.

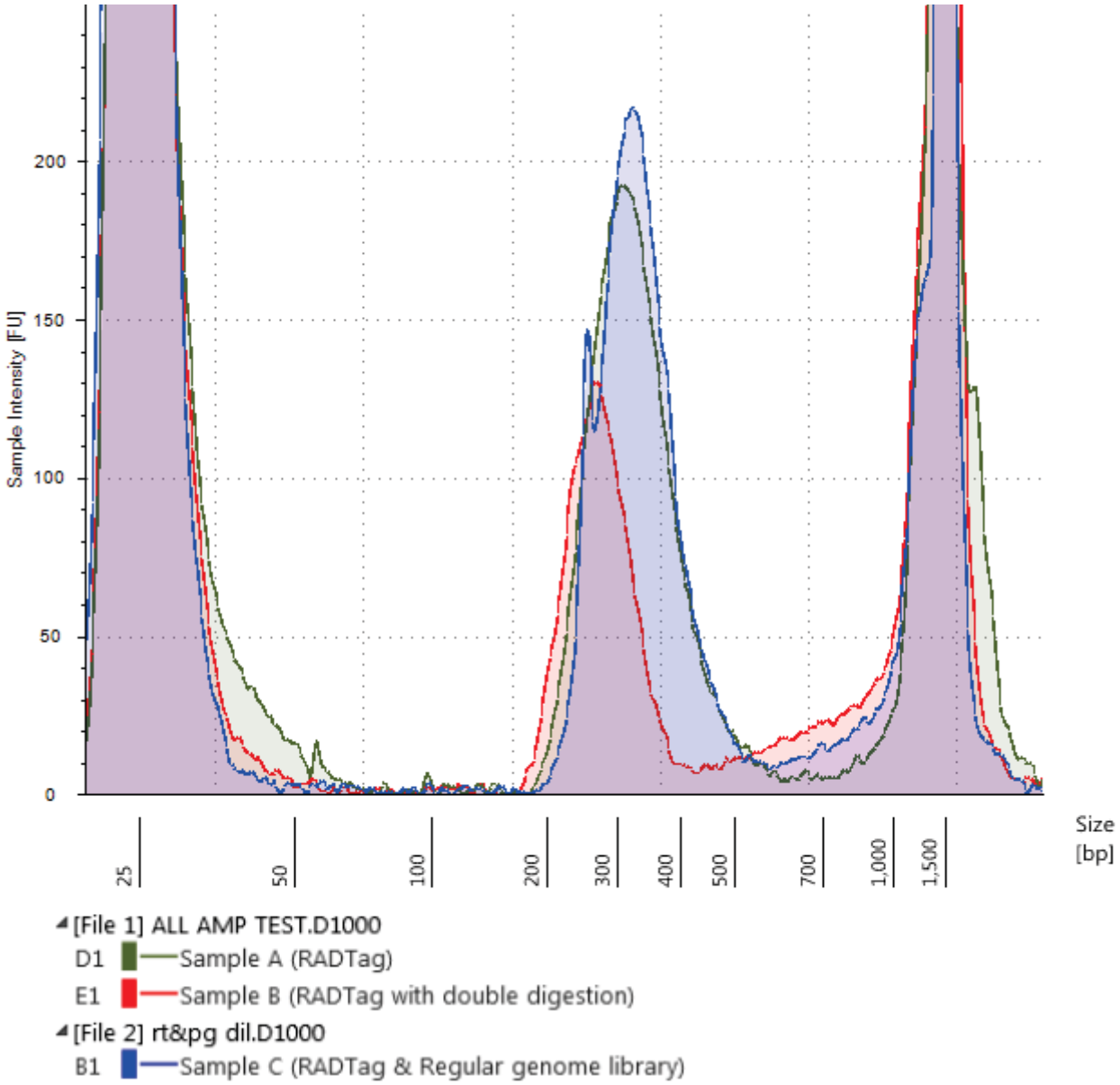


Figure 10: Analysis of Amplified Genomic Samples. An electrophoregram showing the size distribution of amplified samples after size selecting for 310-370 bp using Ampure Beads. An average fragment length of 300, 262 and 326 bp were recorded for Samples A, B and C respectively.

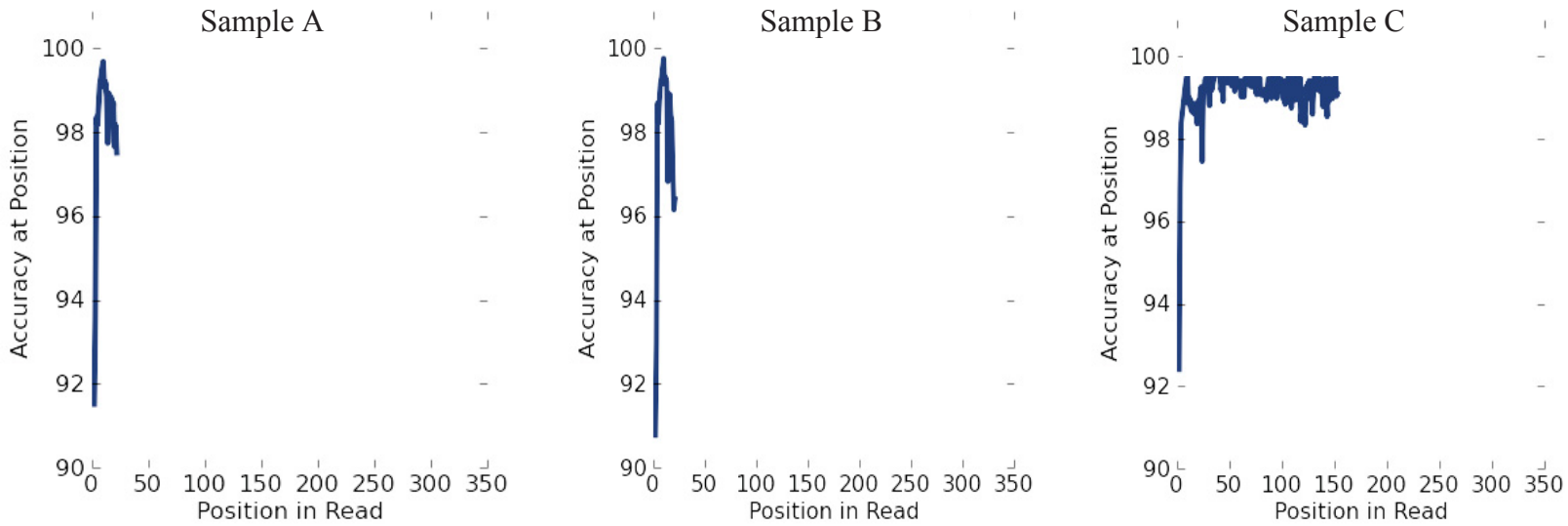


Figure 11: Analysis of Sequenced Mitochondrial DNA contamination. A graph showing the size distribution of mitochondrial DNA sequences present in the nuclear genomic DNA sequences. All sample had mmitochondria separated from the nuclei prior to DNA extract. Sample A and B exhibited mitochondrial DNA sequences of average length < 50bp. Sample C had mitochondria DNA sequences as long as 150bp in its genomic DNA sequences

APPENDIX B

Tables 1 through 7

Table 1: Summary of Major Steps in RADTag Library Prep

Step	Sample A	Sample B	Sample C
Digestion	*Pst 1*	*Pst 1* & *Hae Iii*	*Pst 1*
Fragmentation	Yes	No	No
A-Adaptor Ligation	Yes	Yes	No
End Repair	Yes	Yes	Yes
P-Adaptor Ligation	Yes	Yes	Ligation Of A&P
Size Selection	Yes	Yes	Yes
Linear Amplification	Yes	Yes	Yes
Type Of Output	RADTag	RADTag	RADTag/Partial Genomic

A summary of all reactions involved in the preparation of the RADTag library for all three techniques and the outcome of each technique.

Table 2: Analysis of Samples after Amplification over 8 Cycles

Sample		Concentration		
Sample ID	Library-product	(ng/L)	(nmol/L)	Fragment length (bp)
A	RADTag	2.09	10.7	300
B	RADTag (double digest)	1.13	6.64	262
C	RADTag/partial genomic	25.7	121.9	326

Size selected samples of fragments length ~300 bp were amplified and analyzed using the Aligent 2200 Tapestation Bioanalyzer.

Table 3: Analysis of ISP-Template Conjugate after Emulsion PCR

		Input Concentration		Raw RFU Value		Background RFU Negative Control Tube			
Sample Technique	Sample ID	(pmol/L)	(pg/µL)	AF 488	AF 647	AF 488	AF 647	CF	Templated ISPs%
Ecoli trial	O			1408.5	282.6	40	7	1.5	11.63%
RADTag	A	15	2.9	2317.7	956.8	43.1	7.5	1.5	24.11%
RADTag - Double Digest (*Pst 1- Hae III*)	B	15	2.55	1227	500	39.3	7.8	1.5	23.94%
Partial RADTag/Genomic	C	13	2.7	1653.3	523.4	42	8.2	1.5	18.47%

ISP-Template Conjugates were analyzed using the Qubit Fluorometer and a Conversion Factor (CF) provided by ION Torrent.

Table 4: Analyses of Sequence Runs with Respect to the Template Prep

Samples	Chip wells	Loading	Library ISP %	Polyclonal (%)	Low quality (%)	Final library (%)	Final number of reads
A	6,348,213	4,936,681	99.6	26	9.8	64.1	3,141,915
B	6,348,216	5,049,137	99.2	20.2	6.7	71.5	3,518,587
C	6,348,215	5,204,402	99.6	20.8	4.8	74.4	3,849,445

Table 5: Analyses of Genomic Sequence Fragments for All 3 Samples

Sample	Bases (M)	Load (%)	Mean Length	Median length	Mode length	Total reads	Mitochondrial DNA	Usable reads	Usable gDNA reads (%)
A	552	78	175	182	193	3,141,915	160,070	2,981,845	94.9
B	581	80	162	163	184	3,581,587	176,025	3,405,562	95.1
C	738	82	191	201	213	3,849,445	381,602	3,467,843	90.1

Table 6: Stacks Analysis of Individual Sequenced Samples

	SITES	VARIANT SITES	% VARIANT SITE	OBS HET	EXP HET	OBS HOM	EXT HOM
Sample A	30347080	8975	0.0295745	0.0002957	0.00015	0.9997	0.9997
Sample B	24813497	7126	0.0287182	0.0002872	0.00014	0.99971	0.99971
Sample C	89829	13	0.0144719	0.0001447	7.24E-05	0.99986	0.99986

Sequenced samples were analyzed for SNP using the Stacks Software. The homozygosity and heterozygosity of samples were determined by executing the 'Population' command in the Stacks Software.

Table 7: Stacks Analysis of Combined Sequenced Samples

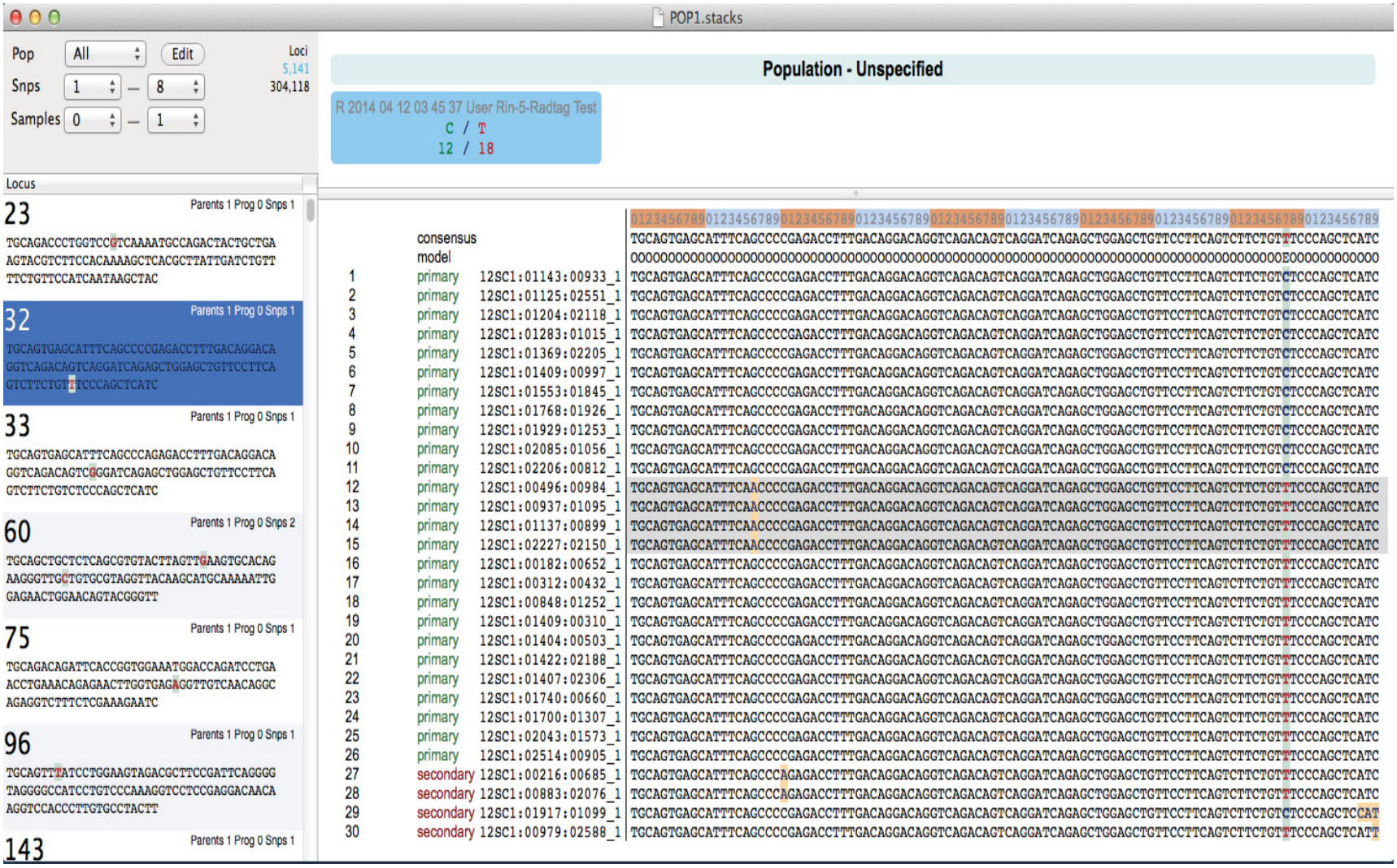
	Fixed Sites	Variant Sites	Variant Sites (%)	Observed Heterozygosity	Expected Heterozygosity	Observed Homozygosity	Expected Homozygosity
Sample A & B	45317175	14400	0.031776	0.0002644	0.00015	0.99974	0.99973
Sample A & C	30432259	9001	0.0295772	0.0002952	0.00015	0.99971	0.99971
Sample B & C	24898577	7146	0.0287004	0.0002865	0.00014	0.99971	0.99971
All Samples	45400755	14426	0.0317748	0.0002641	0.00015	0.99974	0.99973

Sequenced samples were analyzed for SNP using the Stacks Software. The homozygosity and heterozygosity of samples were determined by executing the 'Population' command in the Stacks Software.

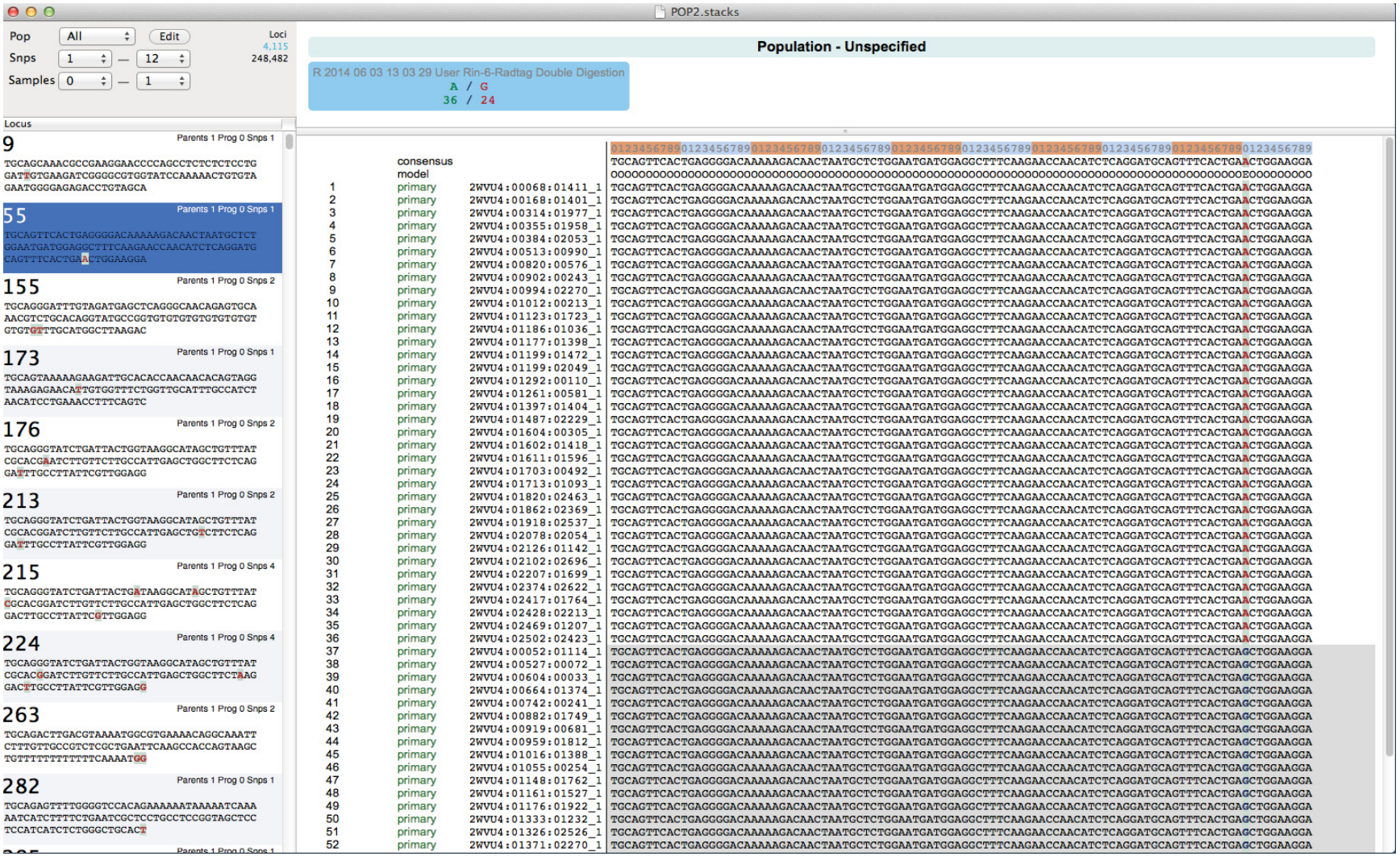
APPENDIX C

Examples of Identified Single Nucleotide Polymorphisms (SNPs)
in Individual sample

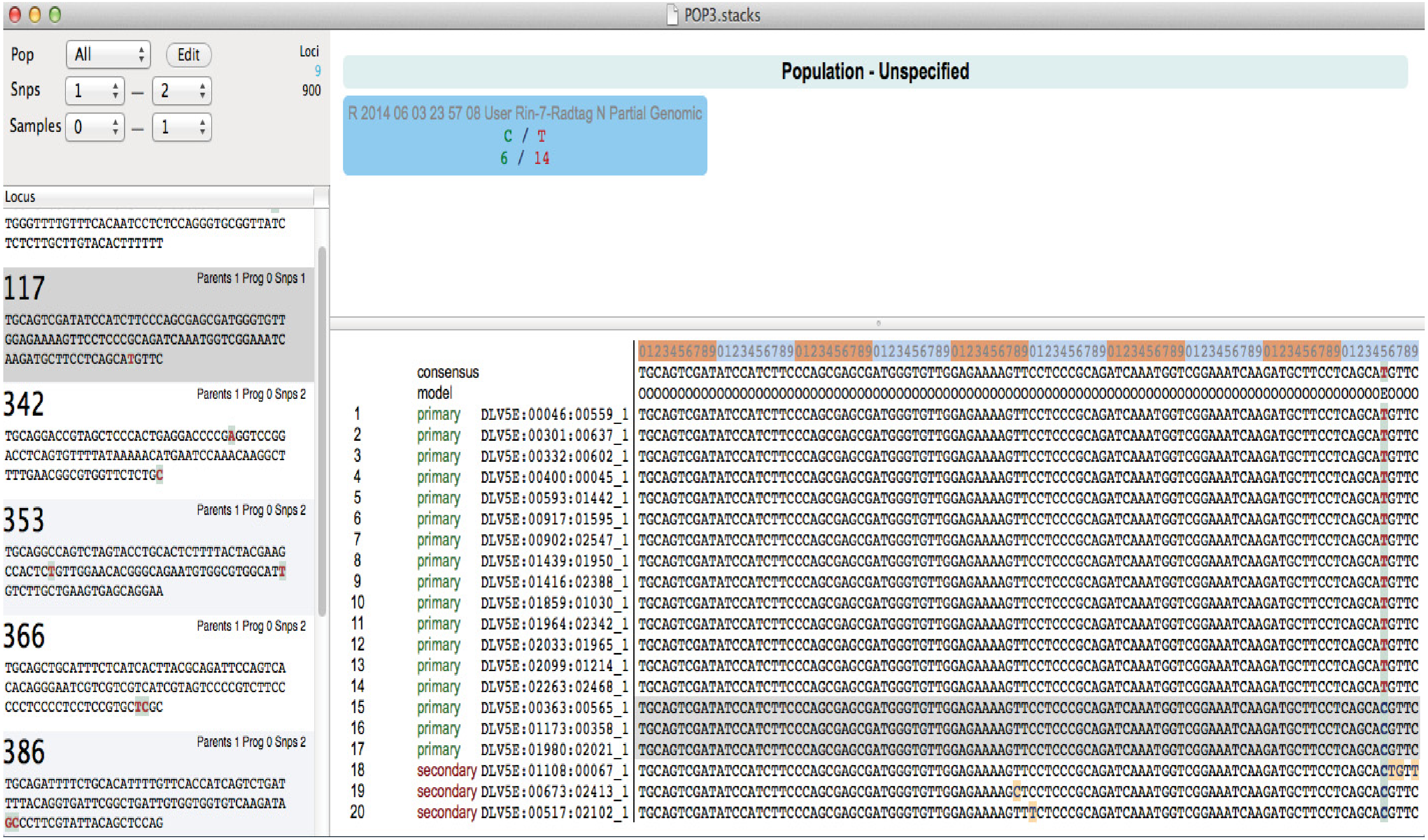
Identified SNPs in Sample A



Identified SNPs in Sample B



Identified SNPs in Sample C



APPENDIX D

Valdosta State University Institutional Animal Care and Use Committee Approval Form

VALDOSTA STATE UNIVERSITY
Building for Our Next Century

October 11, 2012

Dr. John Elder
Department of Biology
Valdosta State University

Dear Dr. Elder:

Your Animal Use Protocol, "Comparison and Identification of *Kreptolebias marmoratus*, *Kryptolebias ocelatus* and Their Hybrid, Gitmo, Using Gene Study" (AUP-00046-2012) has been approved by the Institutional Animal Care and Use Committee (IACUC). This approval is for the period of October 11, 2012, through October 10, 2015. Each year, an annual review and report must be submitted to the IACUC to keep your protocol active. You will be contacted by the Office of Sponsored Programs and Research Administration approximately one month before the annual review and report is due.

Please remember that you must obtain IACUC approval before amending or altering the scope or procedures of the protocol. You are also required to report to the Attending Veterinarian, the IACUC Chair, and the IACUC Administrator any unanticipated problems with the animals which become apparent during the course, or as a result, of the research or teaching activity.

If you have any questions, please contact the IACUC at iacuc@valdosta.edu.

Sincerely,

Barbara Gray

Barbara Gray
IACUC Administrator

cc: Dr. Karla Hull, Institutional Official
Dr. Theresa Grove, IACUC Chair
Dr. Teresa Doscher, Attending Veterinarian
Dr. Robert Gannon, Biology Department Head