2004

The Use Of Pyrosequencing For The Analysis Of Y Chromosome Single Nucleotide Polymorphisms

Jeremy Charles Fletcher
University of Central Florida

THE USE OF PYROSEQUENCING FOR THE ANALYSIS OF Y CHROMOSOME SINGLE NUCLEOTIDE POLYMORPHISMS

by

JEREMY CHARLES FLETCHER
B.S. University of Central Florida, 2000

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Chemistry
in the College of Arts and Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2004

ABSTRACT

The potential value of the Y chromosome for forensic applications has been recognized for some time with the current work dedicated to Short Tandem Repeat analysis and Single Nucleotide Polymorphism (SNP) discovery. This study examined the ability of two different SNP analysis methods to determine if they could be utilized in forensic applications and ultimately be developed into an established system for Y chromosome SNP analysis.

This study examined two principle SNP analysis systems: single base extension and Pyrosequencing. Pyrosequencing was determined to be superior to single base extension, due to the wealth of information provided with sequencing and the flexibility of designing primers for analysis. Using Pyrosequencing, 50 Y chromosome loci were examined and the minimum loci required for maximum diversity for the development of a Y chromosome SNP analysis system were chosen. Thirteen loci were selected based on their ability to discriminate 60 different individuals from three different racial groups into 15 different haplogroups. The Y chromosome SNP analysis system developed utilized nested PCR for the amplification of all 13 loci. Then they were sequenced as groups, ranging from one to three loci, in a single reaction.

The Y chromosome SNP analysis system developed here has the potential for forensic application since it has shown to be successful in the analysis of blood, buccal swabs, semen, and saliva, works with as little as 5 pg of starting DNA material, and will amplify only male DNA in the presence of male/female mixtures in which the female portion of the sample overwhelmed the male portion 30,000 to 1.

ACKNOWLEDGEMENTS

I would like to thank Dr. Ballantyne for his guidance, knowledge, and letting me plow my own path through this project. A special thanks goes out to all the lab rats that put up with my unorthodox personality (you're going to have to set up your own calculations from now on).

I would like to thank my parents for trusting me and allowing me the freedom to do what I felt was best, throughout my entire life.

Most importantly, I would like to thank my wife, Heather, who afforded me the time and space to get my degree and without her, nothing I do in life would have any meaning.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

The Y chromosome is the human male sex-determining chromosome and contains approximately 50 million base pairs [1]. The Y chromosome is divided into three distinct sections: the pseudoautosomals regions, the euchromatin and the heterochromatin. The pseudoautosomals regions make up 5% of the total genomic content and are located at the distal ends of each arm of the Y chromosome (PAR1 and PAR2). The pseudoautosomals regions have 99% homology with the X chromosome and are responsible for the Y chromosomes correct pairing with the X chromosome during meosis [2]. The remaining 95% of the Y chromosome contains approximately 24 Mb of euchromatin and 30 Mb of heterochromatin all within the non-recombining region (NRY). The heterochromatin contains areas of condensed inert genetic material while the euchromatin contains all the functional genes, most having to do with male sexuality and fertility [3]. The NRY as the name implies, does not recombine during meiosis and is therefore passed paternally from father to son without alterations, baring any rare mutation.

The paternal heritage of an unaltered NRY has shown a myriad of possible applications including population migration/ethnicity [4, 5], instances of questioned paternity [6], and forensic applications. Forensically, the Y chromosome is valuable in instances of male/female mixtures, especially when the female portion greatly overwhelms the male portion of the mixture. Such an instance could be achieved with the deposit of semen by an azoospermic or oligospermic male, oral sodomy where only trace amounts of male buccal epithelial cells may be present, or the normal post-coital degradative and sample loss that occurs with the passage of

time [7]. Due to the Y chromosome's non homologous nature with the X chromosome, Y chromosome analysis could eliminate the time-consuming and often inefficient differential extraction procedure. Y chromosome analysis is also forensically useful in instances of disaster identification or missing persons, by requiring the analysis of only a male relative to the person in question. It has even been suggested that this mode of inheritance could enable a particular Y chromosome analysis recovered at a crime scene to be linked with a surname although the confounding effects of non-paternity and multiple independent origins of surnames would need to be taken into account [7].

Analysis of the Y chromosome can be accomplished through examination of one of the two major types of genetic variants: tandem repeats (such as short tandem repeats (STRs)) or single nucleotide polymorphisms (SNPs). SNPs are the smallest unit of mutation which allows for their analysis in degraded samples, but have a low number of alleles present at each mutation point compared with STRs. To overcome this drawback, a larger number of SNPs need to be examined to get a discriminating potential similar to a minimal number of STRs. This is not a problem due to the abundance of SNPs located within the genome (over 1.42 million, and more than 4000 on the Y chromosome alone [8]) and the potential for automation with numerous analysis techniques.

With the emergence of SNPs and their overwhelming number, many methods have been established for their identification. Various methods utilize allele specific oligonucleotide (ASO) hybridization to differentiate between allele variants. A drawback to this technique is the propensity for mispriming which can cause false positive results [9]. Another common SNP

analysis technique is single base extension. This technique utilizes a primer that binds to the template DNA one base away from the SNP of interest. DNA polymerase is then used to extend the primer one nucleotide, complimentary to the SNP located in the template. Once the primer is extended multiple analysis techniques are used to determine which of the four dNTPs (Adenosine, Cytosine, Guanine, or Thiamine) was incorporated.

One method of single base extension currently available is the ABI Prism SNaPshot ddNTP Primer Extension. SNaPshot is based on the dideoxy single nucleotide extension of an unlabeled oligonucleotide primer. An oligonucleotide primer binds to a complimentary DNA template one nucleotide (5') to an SNP of interest. In the presence of Amplitaq DNA polymerase and fluorescently labeled ddNTPs, the primer is extended one nucleotide by the addition of one of the fluorescently labeled ddNTPs [10]. To determine which base was incorporated and subsequently to identify the SNP of interest the extended products are run through an ABI Prism 310 Genetic Analyzer. The base that was incorporated can be determined by the florescent signal observed, because each nucleotide is labeled with a different florescent dye (Table 1). For example, a green florescence signal corresponding to the elution time when the primer plus one sized nucleotide exits the column indicates that the oligonuclotide primer was extended using an Adenosine nucleotide. Therefore the template DNA had a Thymine nucleotide 3' to the primer binding site. This technique can be problematic if the DNA sequence adjacent to the SNP is not conducive for primer design or binding. Primers which have complimentary regions between their 3' end and internal sequences can form primer dimmers or hairpins and ultimately lead to dNTP incorporation and false positives [11].

Table 1

Dye Color Arrangements

ddNTP	Dye	Dye Color
A	dR6G	Green
C	dTAMARA	Yellow
G	dR110	Blue
T	dROX	Red

The most informative SNP analysis technique is sequencing. Sequencing the DNA around a SNP provides sequence contex for accurate genotype conformation. Since sequencing the DNA can begin at any site there are more options in choice for the sequencing primer, while also leading to easier elimination of false positives from primer interactions. Conventional sequencing utilizes dideoxy chain termination methodology, but has limitations in both throughput and cost for SNP analysis [12]. An alternate method of DNA sequencing is pyrosequencing. Pyrosequencing is a sequencing by synthesis technique which provides for an accurate and consistent analysis of DNA sequences [13]. This methodology sequences DNA by monitoring the release of pyrophosphate during the incorporation of known nucleotides in DNA synthesis.

Pyrosequencing begins by hybridizing a sequencing primer to a PCR amplified ssDNA template. This hybridized template/primer combination is then mixed with an enzyme cocktail containing DNA polymerase, ATP sulfurylase, luciferase, and apyrase, as well as a substrate mix of adenosine 5' phosphosulfate (APS) and luciferin. Once the template/primer, enzymes and

substrates are mixed, one of four deoxynucleotide triphosphates (dNTP) is added to the reaction. If the added nucleotide is complimentary to the template nucleotide, 3' to the extension primer, then DNA polymerase will incorporate the dNTP into the extension primer DNA strand. This incorporation causes the release of pyrophosphate (PPi) in amounts proportional to the number of incorporated nucleotides (Figure 1a). If the added nucleotide is not complimentary to the template nucleotide 3' to the extension primer then DNA polymerase will not incorporate the dNTP and PPi will not be released [13].

The enzyme ATP sulfurylase converts the released PPi to adenosine triphosphate (ATP) by combining PPi and adenosine 5' phosphosulfate. The amount of ATP produced is proportional to the amount of PPi created by DNA polymerase. The enzyme luciferase uses the ATP produced by ATP sulfurylase to convert luciferin to oxyluciferin, which generates photons of light. The amount of light produced by luciferase is directly proportional to the amount of ATP produced. The light produced is detected with a charged coupled device (CCD) camera and is depicted as a peak in the pyrogram (Figure 1b). The height of the peak is proportional to the amount of light produced and therefore proportional to the number of nucleotides incorporated[13].

While the enzyme cascade producing the luciferase-catalyzed light is in progress, apyrase is continuously degrading any ATP or unincorporated dNTPs. After a few seconds, all the ATP and dNTPs are degraded and the reaction is reset for the addition of the next nucleotide. Due to the dNTPs being added in a known sequence one at a time, the created pyrogram is representative of the complimentary strand of the PCR amplified DNA template (Figure 1c) [13].

Since sequencing is the preferred method of SNP evaluation (in terms of assuring accurate allele calls) pyrosequencing was evaluated to determine if it can stand up to some of the challenges required for Y-chromosome forensic use. Special attention was placed on the handling of minimal amounts of starting sample, multiple DNA sources, and mixtures of male DNA in overwhelming amounts of female DNA.

It is crucial for any DNA based technology to be able to handle minimal amounts of starting DNA, since DNA evidence is routinely found in limited quantities. The ability to handle multiple DNA sources is required since all possible sources of DNA can be found and must be able to be used for analysis in forensic science. Male female DNA mixtures are commonly found in sexual assault (semen or saliva deposited in or on the female victim) and can also be found in other assaults when blood and other body fluids are deposited onto one another. In cases like these, especially in sexual assault, the female portion of the mixture can greatly overwhelm the male portion of the mixture. Therefore in traditional autosomal analysis of sexual assaults the male/female mixtures are divided into a male, sperm fraction, and a female, epithelial fraction for analysis. This can be problematic, since sample containing little male DNA can be routinely lost during the differential extraction procedure to separate the male and female DNA. Y chromosome analysis eliminates this problem since females do not have a Y chromosome, the samples can be analyzed without separating or losing the male DNA.

In order to minimize the amount of starting DNA required for analysis, limit the number of biotinylated templates in the pyrosequencing mixture, maximize our haplogroup diversity, all while maintaining proper Y-chromosome SNP allele calling a nested PCR system was

developed. This system amplifies 13 Y-chromosome SNP loci, approximately 300 – 500 base pairs in size in the first of two rounds of amplification. The second round of amplification takes one to three of the starting 13 loci and amplifies them using the biotinylated primers necessary for Pyrosequencing (amplicons approximately 100 - 300 base pairs in size). This second biotinylated amplified product is then used for pyrosequencing the 13 Y-chromosome loci in three multiplexes (PMP1, PMP2, PMP3) and five singleplexes (M3, M74, M172, M24, and P25) (Figure 2).

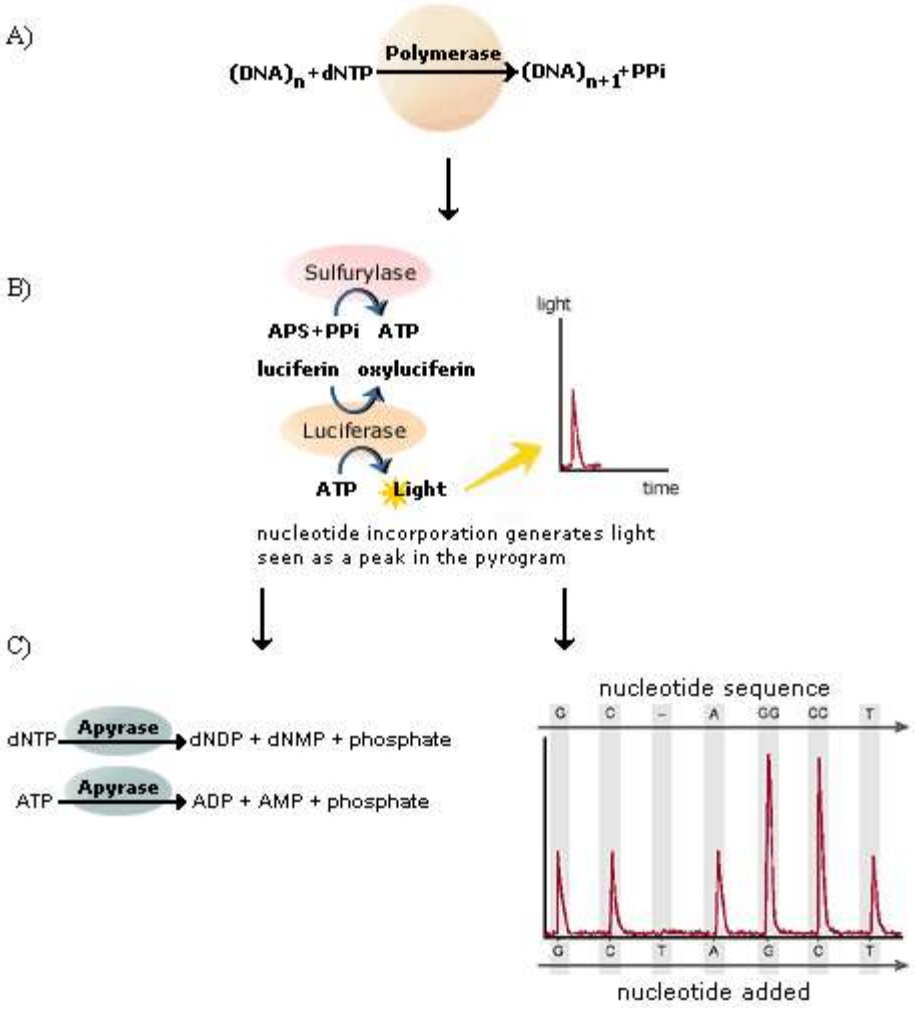


Figure 1. Priciples of Pyrosequencing. A) dNTP is incorporated onto extending sequencing primer, causing release of pyrophosphate B) Sulfurylase convert pyrophosphate to ATP, Luciferase converts ATP to light, and light is detected with CCD camera C) Apyrase degrades unincorporated dNTPs and unused ATP to reset reaction and allow for the next nucleotide to be added while a pyrogram is being produced [13]

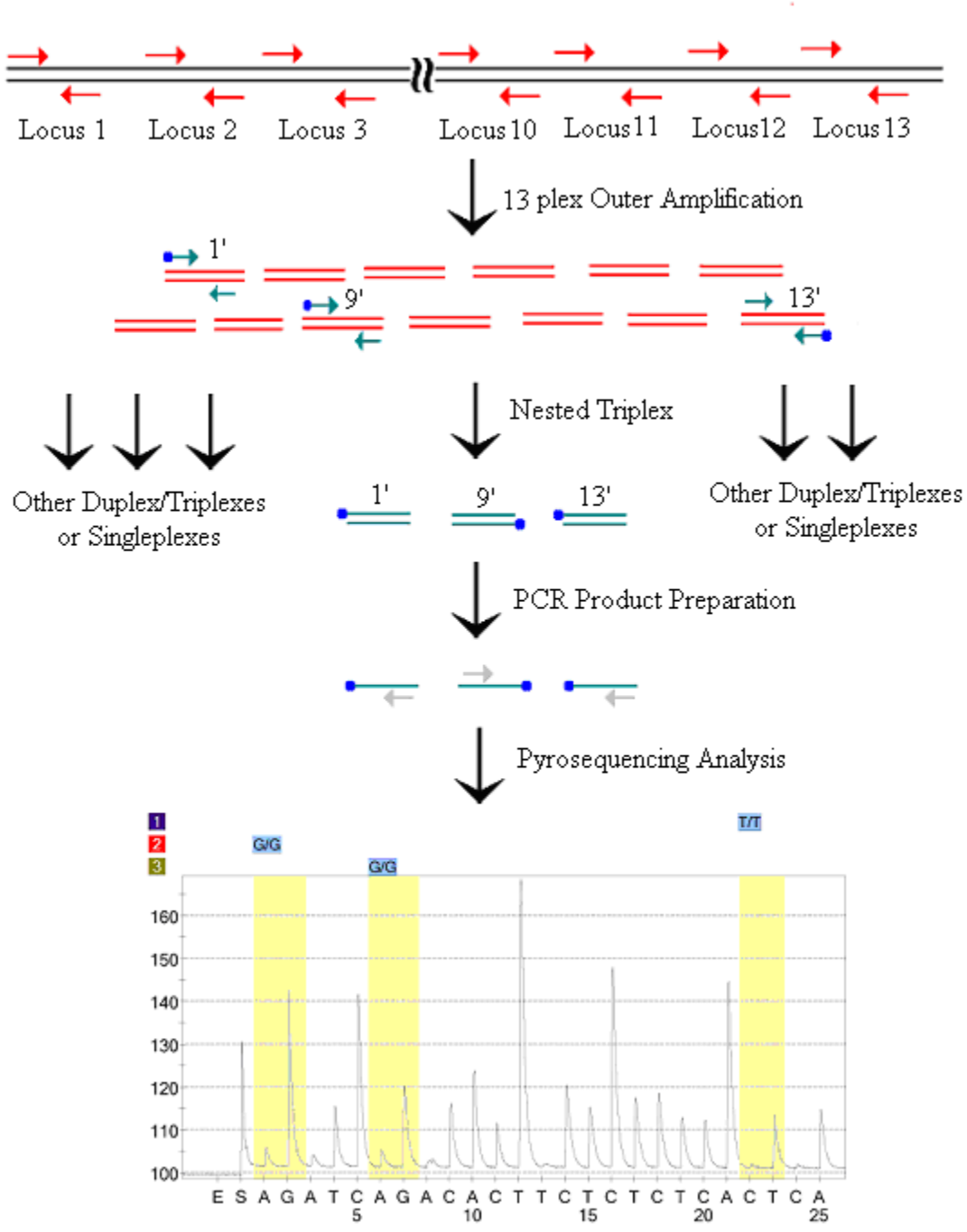


Figure 2. System Overview. Overview of Y-chromosome SNP analysis system from amplification of original starting DNA with 13 plex primer set to analysis of Pyrosequencing Multiplexes.

CHAPTER 2. METHODS

2.1 Preparation of Body Fluid Stains

Body fluids were collected from volunteers using procedures approved by the University of Central Florida's Institutional Review Board. Buccal samples were collected from donors using sterile swabs by swabbing the inside of the donor's mouth. Population samples for gene diversity and haplogroupings were obtained from local Orlando residents (buccal swabs), the Virginia Division of Forensic Science (bloodstains), and the Orange County Sheriff's-Coroner's office in Santa Ana, CA (DNA extracts). All samples were stored at -20^{o}C until needed.

2.2 DNA Isolation and Purification of Buccal Swabs, Saliva and Semen

DNA was extracted from the buccal swabs and semen using a standard phenol:chloroform method [14]. Stains or swabs were cut into small pieces and placed into a Spin-Ease tube (Gibco-BRL, Grand Island NY). The tubes containing buccal samples and saliva were incubated overnight in a 56^{o}C water bath using 400 µl DNA Extraction Buffer (100mM NaCl, 10mM Tris-HCl, pH 8.0, 25mM EDTA, 0.5% SDS), and 0.1mg/mL Proteinase K. The tubes containing semen samples were incubated overnight in a 56^{o}C water bath using 400 µl DNA Extraction Buffer, 0.1mg/mL Proteinase K, and 40µL dithiotreitol (DTT). After the overnight incubation, all samples were placed into a Spin-Ease basket, the basket inserted back into the original tube, and the samples centrifuged at 14,000g for 5 minutes to remove the absorbed fluid from the swab material. A volume of phenol/chloroform/isoamyl alcohol equal to

the volume of the crude extract was added and vigorously intermixed by shaking The aqueous layer, containing the DNA, was removed. Precipitation of the DNA was accomplished by the addition of cold absolute ethanol (two and a half times the volume of the aqueous layer extract) and allowed to precipitate overnight at –20°C. The DNA was pelleted by centrifugation, washed twice using 70% ethanol and re-solubilized with 100 µl of TE^{-4} (10 mM Tris-HCl, 0.1 mM EDTA, pH 7.5) overnight at 56°C.

2.3 DNA Isolation and Purification of Dried Blood Samples

The dried blood stains were incubated overnight at 56°C in 400 µl of DNA extraction buffer (100mM NaCl, 10 mM Tris-HCl, 25 mM EDTA, 0.5% SDS) and 0.1mg/mL Proteinase K. The swab pieces were placed into a spin ease basket and centrifuged at 14,000g for 5 minutes. An equal volume of phenol/chloroform/isoamyl alcohol was added to the crude extract. The aqueous phase extracts containing the DNA were purified using Centricon 100™ concentrators (Millipore, Bedford, MA) according to the manufacturer's instructions.

2.4 Gel Electrophoresis

2.4.1 DNA Quantitation Gel

Extracted DNA was quantitated using a 1% agarose gel containing 0.01% Ethidium Bromide. The quantitation gel was made by adding 30 µL of 5 mg/ml Ethidium Bromide (EtBr)

solution to 300 mL of 1X TAE (Tris-Acetate EDTA) Buffer. 30 mL of the TAE/EtBr solution was combined with 0.30 g of molecular biology grade agarose (Fisher, Suwanee, GA), and heated to a boil in a microwave. The solution volume was measured to determine the quantity of water lost due to boiling. Deionized water was used to bring the final volume up to the original 30 mL. The 30 mL TAE/EtBr/ Agarose solution was poured into a Minigel gel-casting tray containing two 12-tooth combs and allowed to harden at room temperature. The remaining 270 mL of the TAE/ EtBr solution was poured into the Minigel Electrophoresis chamber and the hardened gel was also placed in the Minigel electrophoresis chamber. The gel was then lifted and tilted to allow for any bubbles in or around the gel casting tray to be removed. A 5 µL aliquot of the DNA samples to be tested were combined with 1 µL of 6X loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol, 40% sucrose), vortexed then spun. The upper lanes of the gel were loaded with 6 µL of the ten DNA standards (5ng, 15ng, 30ng, 60ng, 100ng, 125ng, 200ng, 375ng, and 500ng per 6 µL volumes). The lower lanes of the gel were loaded with the 6 µL sample/loading buffer mixture. Gels were run for 12-15 minutes at 200V. The DNA bands were visualized using a UV transilluminator, and a visual record taken either with a Polaroid camera or the Omega 10™ photodocumentation system. Calculating the concentration of DNA in each sample was made by visual comparisons to the known standards.

2.4.2 PCR Product Gel

PCR Product gels were made to the same specifications as the gels used for DNA Quantitation except that only 1 comb was used, located at the top of the gel. These gels were then run for 25 minutes at 200 V, and visualized the same as the DNA Quantitation gels.

2.5 Primers

2.5.1 Initial PCR Primers

All initial PCR primers used were designed using one of two sources: Oligo 6 Primer Analysis Software (Lifesciences Software Resource, Long Lake, MN) or Primer3 [15]. Using Oligo 6, primers were selected which had the highest P.E. scores relative to the other primer options that produced a PCR product preferably less than 300 base pairs. Using Primer3, primers were selected that had minimal 3' self-complementarity of the primer or oligo ("taken as a measure of its tendency to form a primer-dimer with itself") or any self-complementarity ("taken as a measure of its tendency to anneal to itself or form secondary structure") [15] which had PCR product sizes preferably less than 300 base pairs.

If the primer pairs were to be used with pyrosequencing, the final PCR product oligonucleotide sequence was then entered into pyrosequencing's SNP Primer Design Software v 1.01 [16] and analyzed for sequencing primers. During the analysis of sequencing primers, the SNP Primer Design Software would determine if the PCR product template would form a hairpin loop with itself. If this were the case, then alternate forward, reverse or both forward and reverse

primers would be redesigned using the same methods listed above. Determining which primer (forward or reverse) would be labeled with the biotin was determined by examining the number of different options for the forward and reverse analysis, and the simplicity with which allele calls would be made. For example, if five of the six possible extension primers in the forward direction caused primer dimmers and could not be used, then the reverse primer would be labeled. If the analysis in the forward direction yielded a SNP within a homopolymer then the reverse primer would be labeled.

2.5.2 Extension Primers

Extension Primers for SNaPshot were designed by taking 20 bases complimentary to either template, 5' to the SNP site. Extension primers for Pyrosequencing were designed using SNP Primer Design Software v 1.01. Primers were selected based on the final score provided by the design software. The SNP Primer Design Software evaluates scores based on seven criteria: Melting Temperature (reducing the score for melting temperatures below 50˚C), Mispriming Analysis (reducing the score for alternate priming sites for the sequencing primer), Duplex Formation Analysis (reducing the score for finding the existence of extension primer dimmers), Hairpin Loop Analysis (reducing the score for hairpin extension primer hairpin loops), Repeat Bases at SNP Analysis (reducing the score for homopolymers at the site of SNP) Primer length (reducing the score for primers over 18 nucleotides due to their high price) and Template Loop Analysis (reducing the score for oligonucleotide templates looping back onto itself). The total score was calculated using a complicated arithmetic formula using all the scores from each of the

various scoring criteria. The highest scoring primer with 19 or more nucleotides was selected for pyrosequencing analysis when many primers were acceptable. When certain extension primers were needed for multiplex analysis those primers were chosen regardless of final score.

2.5.3 External PCR Primers

External PCR primers used in nested PCR were designed using Primer3. Primers were selected which had relatively low 3' or any self-complementarity, and had as few base pairs between the newly designed external primer pairs and the previously designed internal PCR primers. Preferentially the PCR product size would be less than 450 base pairs.

2.6 PCR Conditions

2.6.1 SNaPshot PCR Condition

Amplification of genomic DNA for SNaPshot analysis utilizes a 50 μL reaction containing: 0.5 – 3 ng of DNA, 1.6 μM forward and reverse primers, 1X PCR Buffer I (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl), 1 mM dNTP, 2mg/ml BSA, and 0.5 units of Amplitaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA). Cycling conditions for amplification are: (1) 95˚C for 11 min (2) 32 cycles of 96˚C for 30 sec, 56˚C for 60 sec, and 72˚C for 60 sec (3) 72˚C for 10 min (4) 4˚C for ∞.

2.6.2 Pyrosequencing PCR Conditions (Direct)

Amplification of genomic DNA for direct Pyrosequencing analysis utilizes a 50 µL standard reaction established by pyrosequencing containing: 0.5 – 2 ng of DNA, 2.0 µM forward and reverse primers, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 2mM MgCl, 500 µM dNTP, 2mg/ml BSA, and 0.3 units of Amplitaq Gold DNA Polymerase. Cycling conditions for amplification are: (1) 95˚C for 5 min (2) 45 cycles of 95˚C for 15 sec, 50˚C for 30 sec, and 72˚C for 15 sec (3) 72˚C for 5 min (4) 4˚C for ∞.

2.6.3 Pyrosequencing PCR Conditions (Nested)

Amplification of genomic DNA for Nested PCR and ultimate Pyrosequencing analysis utilizes two different reactions for two different methodologies (multiplex and singleplex), and are as follows: *Single Locus Amplification for Outer Primer Sets (50 µL reaction):* 0.5 – 1 ng of genomic DNA, 1.2 µM forward and reverse primer, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 1.5 mM MgCl, 500 µM dNTP, 2mg/ml BSA, and 0.3 units of Amplitaq Gold DNA Polymerase. *Single Locus Amplification Inner Primer Sets (50 µL reaction):* 1 µL of outer PCR product, 2.0 µM forward and reverse primers, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 1.5 mM MgCl, 500 µM dNTP, 2mg/ml BSA, and 0.3 units of Amplitaq Gold DNA Polymerase. *Multiple Locus Amplification for Outer Primer Sets (50 µL reaction):* 0.5 – 1 ng of genomic DNA, 1.2 µM forward and reverse primers, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 1.5 mM MgCl, 1 mM dNTP, 2mg/ml BSA, and 0.5 units of Amplitaq Gold DNA Polymerase. *Multiple Locus Amplification for Inner Primer Sets (25 µL reaction):* 0.5 µL

of outer PCR product, 0.48 - 8 μM forward and reverse primers, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 1.5 mM MgCl, 1 mM dNTP, 2mg/ml BSA, and 0.5 units of Amplitaq Gold DNA Polymerase. Cycling conditions for all nested amplifications were: (1) 95˚C for 10 min (2) 30 cycles of 95˚C for 30 sec, 54˚C for 60 sec, and 72˚C for 30 sec (3) 72˚C for 10 min (4) 4˚C for ∞.

2.7 SNaPshot Sample Preparation

Template Preparation: 4 μL SNaPshot PCR products were combined with 2 units SAP (United States Biochemical, Clevland, OH) and 2 units **ExoSAP-IT®** **(**United States Biochemical) and incubated at 37˚C for 60 min followed by heating at 72˚C for 15 min. *Reaction Preparation:* 5 μL SNaPshot Ready Reaction Premix (Applied Biosystems), 1 μL Extension Primer (20 μM), 1 μL prepared template, and 3 μL deionized water were combined. The 10 μL Reaction mixture was then extended using a thermocycler and the following conditions: (1) 25 cycles of 96˚C for 10 sec, 50˚C for 5 sec, and 60˚C for 30 sec (2) 4˚C for ∞. *Post-Extention Treatment:* The extended reaction mixture was incubated with 0.5 units of SAP for 15 min. at 72˚C.

2.8 SNaPshot Sample Analysis

To a 0.5 mL 310 Genetic Analyzer tube (Applied Biosystem) 10 μL of deionized formamide was mixed with 1 μL of post treatment sample. The Formamide/Sample mixture was

vortexed and spun before being incubated at 95˚C for 5 min. The Formamide/Sample mixture was placed into an Applied Biosystem 310 Genetic Analyzer and run using the GS POP4 (1mL) E module with a 10 sec injection time, 15 kV electrophoresis and EP voltage, 24 min collection time, 60˚C plate temperature, 150 sec syringe pump time, and a 120 sec pre-injection EP time.

2.9 Pyrosequencing Sample Preparation

Procedure for template preparation was a modification from the original instrument procedures included with the PSQ 96MA instrument [17]. All reagents were mixed thoroughly and utilized at room temperature. 20 µL of Biotin labeled PCR product was placed into a 96 well filter plate. 24 µL of 1:5 mix of Steptavidin Sepharose beads/Binding buffer (10 mM Tris-HCl, 2 M NaCl, 1 mM EDTA, 0.1% Tween 20) was added to the PCR product. PCR product/Steptavidin Sepharose beads/Binding buffer mixture was mixed on a shaker at approximately 1500 rpms for a minimum of 10 minutes. The shaken mixture was placed on a vacuum manifold to remove all the liquid in the filter plate, leaving the biotin labeled PCR product and Steptavidin beads. 50 µL of Denaturation solution (0.2 M NaOH) was added to the filter plate and allowed to incubate for a minimum of 60 sec. Vacuum was applied to remove the denaturation solution, and the filter plate was rinsed twice with approximately 500 µL of Washing buffer (10 mM Tris-Acetate, pH 7.6), removing the wash buffer with vacuum between rinses. 50 µL of Annealing buffer (20 mM Tris-Acetate, 2 mM Mg-Acetate pH 7.6) was then added to the filter plate. The Annealing buffer and Streptavidin bead/Biotin template were then removed from the filter plate using a micropipetor making sure to resuspend the Streptavidin

beads by pipeting up and down 4-5 times before removal. The Annealing buffer/Streptavidin bead/Biotin template was added to a PSQ 96 Plate Low after 3 μL of 5 μM sequencing primer was added to the appropriate well. The PSQ 96 Plate Low was then heated at 80° C for 2 minutes using the PSQ 96 Sample Prep Thermoplate Low. After 2 minutes the PSQ 96 Sample Plate Low was removed from the thermoplate, allowed to cool to room temperature and was placed within the PSQ 96MA instrument.

2.10 Pyrosequencing Dispensation Order Design

2.10.1 Singleplexes

Singleplexes (analysis of only one SNP in one reaction at one time) were designed using the PSQ 96MA 2.1 Software. Using the SNP/Sequences submenu, the sequence surrounding the SNP of interest, corresponding to the labeled template strand of DNA, was inputted and all possible extension primers were selected. When a possible extension primer is selected the PSQ 96MA 2.1 Software provides the sequence of nucleotides to be analyzed from the end of the extension primer to end of the inputted data, complimentary to the template strand of DNA. To produce the order of nucleotides dispensed for SNP analysis the SNP/Simplex entry submenu was used. The extension primer of choice was brought into a blank entry form and a dispensation order was created by the PSQ 96MA 2.1 Software. The dispensation order includes a blank nucleotide to begin the dispensation, the nucleotides needed for analysis (including at least 3 nonvariable peaks within close proximity to the SNP variable region), a blank nucleotide

following the SNP variable region, and if any homopolymers are over 8 nucleotides long then that nucleotide was dispensed twice to ensure all the nucleotides are incorporated.

2.10.2 Multiplexes

Multiplexes (analysis of two or three SNPs in one reaction at one time) were designed in a similar manner to the Singleplexes using the PSQ 96MA 2.1 Software. Using the SNP/Sequences submenu, the sequences surrounding the SNPs of interest corresponding to the labeled template strands of DNA were inputted and all possible extension primers were selected. To produce the order of nucleotides added for SNP analysis the SNP/Multiplex entry submenu was used. Extension primers from up to three different SNPs could be brought into a blank entry form. The PSQ 96MA 2.1 Software creates a dispensation order that best fits the three sequences to analyze. PSQ 96MA 2.1 Software attempts to make the first nucleotide a blank dispensation, then the order of nucleotide dispensation is created to analyze all of the SNPs so as to not overlap a variable site with a constant region from another template. The Software also attempts to provide at least one non variable nucleotide in close proximity to the SNP variable site for each template, to not have a SNP variable region as part of a homopolymer, and to provide a blank dispensation after each variable region. If any of these conditions were not met, PSQ 96MA 2.1 Software would provide a warning, and, either alternate extension primers were chosen to alter the sequences to be analyzed, or entirely new SNPs were selected to be combined in the multiplex. Figure 3 shows an example of how changing the extension primer for one of the two SNPs in a Duplex can alter the ability to provide unambiguous SNP determinations. Figure 3a

shows PMP3 using M170E(2) and M9E(4). If these sequence primers were used with one another, the dispensation pattern would produce an overlap in reference sequence from M170 and the variable region from M9. Overlapping of variable regions can lead to uncertain allele calls using the computer algorithms. To correct this problem, multiple extension primer combinations must be tried until the dispensation order produced non-overlapping variable regions. Figure 3b shows how changing M170E(2) to M170E produces a dispensation order yielding non overlapping variable regions, therefore allowing for easier allele calls during analysis. This same approach was used in eliminating as many warnings as possible before analysis. In some instances, warnings were not able to be eliminated due to the existing sequences surrounding SNP, and had to be accepted.

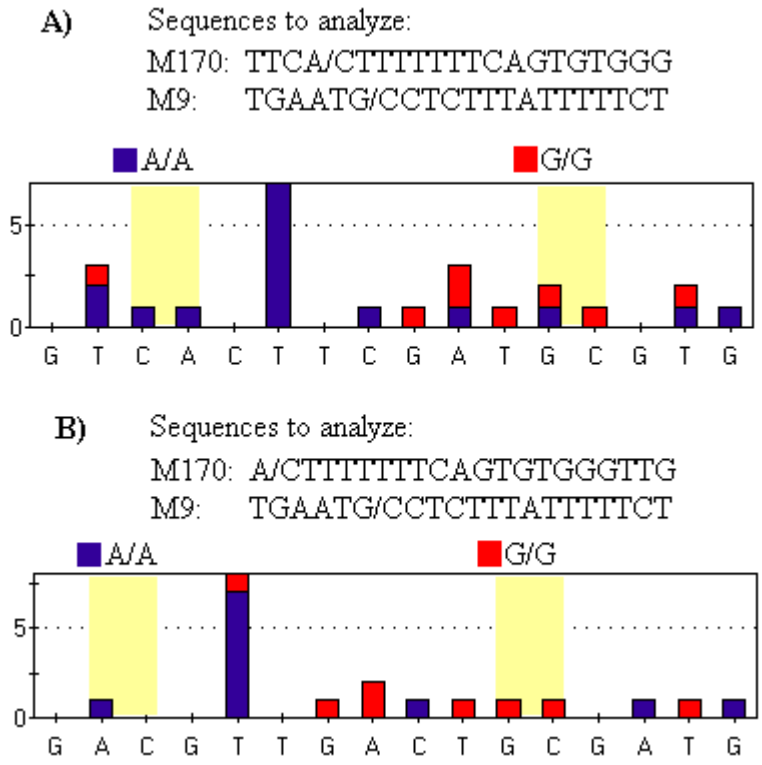


Figure 3. Change in Extension Primer Choice. A) Dispensation order produced when PMP3 is analyzed using M170E(2) and M9E(4), leading to overlapping sequences from M170 reference sequence and M9 variable region (G nucleotide in highlighted region of chart). B) Dispensation order produced when PMP3 is analyzed using M170E and M9E(4), eliminating the overlapping sequences seen in (A).

2.11 Pyrosequencing SNP Analysis

The pyrosequencing reaction takes place using the PSQ 96 SNP Reagent Cartridge filled with the appropriate amount of each nucleotide (50 µL + 0.22 µL x the number of dispensations of the nucleotide over all samples in the run), Enzyme (50 µL + 5.5 µL x the number of wells) and Substrate (50 µL + 5.5 µL x the number of wells). The PSQ 96MA instrument is run using the "instrument parameters code 0001" (reagent priming time 100 ms, nucleotide priming time

30 ms, enzyme pulse time 90 ms, substrate pulse time 88 ms, reagent pressure 400 mbar, nucleotide pulse time 8.5 ms, nucleotide pressure 650 mbar, block temperature 28˚C, block mixer frequency 35 Hz, and cycle time 65 s), and analyzed using Pyrosequencing 96MA SNP Analysis Software.

2.12 Forensic Value Analysis

2.12.1 Polymorphism/Y-specific Analysis

To determine if the SNP of interest was polymorphic and Y-specific, the SNP was examined in 18 African American individuals (2 from Orlando local residents and 16 from Virginia Division of Forensic Science), 20 Caucasian individuals (14 from Orlando local residents and 6 from Virginia Division of Forensic Science), 21 Hispanic individuals (1 Hispanic originating from the South East United States collected from a local Orlando resident and 20 Hispanics originating from the South West United States collected from California) and 2 controls (male-9948, and female-9947a). Each individual tested was analyzed using a singleplex system to ensure the proper allele call.

2.12.2 Various DNA Sources

To determine if the pyrosequencing method of analysis could handle different DNA sources, samples of semen, blood, saliva, and buccal swabs were extracted from the same individual. Each of the samples was amplified using 1ng of sample and analyzed using 13plex

outer amplification, then PMP1, PMP2, PMP3 multiplex analysis and M3, M74, M172, M214, and P25 singlplex analysis.

2.12.3 Minimal Starting DNA

To determine the minimal starting DNA required for analysis of the Nested Y Chromosome SNP system, extracted DNA at 1 ng/µL was diluted to various concentraiotions from 500 pg/µL to 1 pg/µL. These samples were amplified using the outer 13-plex primers, then this product was utilized for the nested amplification of PMP1, PMP2, PMP3, M3, M74 M172, M214, and P25. These products were pyrosequenced and analyzed for the lowest amount of starting DNA that provided an unambiguous allele call for all 13 loci.

2.12.4 Male/Female Mixture Analysis

To determine the power of the nested system to work in male/female mixtures, 1 ng of male DNA was mixed with 8.2 µg, 5.0µg, 1.0µg, 500 ng, and 100ng of female DNA. 300ng of female DNA was also mixed with 25pg, 10pg, 5pg, and 1pg of male DNA. These samples were amplified using the outer 13-plex primers, then this product was utilized for the nested amplification of PMP1, PMP2, PMP3, M3, M74 M172, M214, and P25. These products were pyrosequenced and analyzed for the lowest amount of starting DNA that provided an unambiguous allele call for all 13 loci.

CHAPTER 3. RESULTS

3.1 Loci and Primers

Each loci examined was identified through one of two sources: reference information found in background literature searches, or examinations of listed Y chromosome SNPs in the dbSNP (http://www.ncbi.nlm.nih.gov/SNP) and the SNP consortium (http://snp.cshl.org) on-line databases. Table 2 lists all the Y chromosome loci examined with either SNaPshot or Pyrosequencing. Y chromosome loci were chosen based on referenced sequence variation in Caucasian and African American populations or their position within the Y chromosome Consortium's Y haplotype tree [18] (Figure 4). Table 3 lists the PCR primers that were designed and shown to produce acceptable PCR products for the Y chromosome SNPs examined. Table 4 lists all the extension primers that were designed for use in both SNaPshot and Pyrosequencing. Table 5 lists the PCR primers designed and used for amplification of outer PCR products with nested PCR and the 13plex. 13plex loci were chosen based on polymorphism data. All loci that had been evaluated for forensic value through polymorphic analysis at 13plex design initiation, that had been determined to be polymorphic and produced the most Y chromosome haplogroups, without overlapping any loci data were chosen for inclusion in the 13plex. All primers were designed as described in chapter 2.

Table 2

Loci Reference List

	Position (BLAST)			
Name	Access #	BP #	SNP	References
M2	AC011302	24878	G/A	[1, 5, 18-20]
M3	AC007034	95124	G/A	[1, 5, 18-23]
M6	AC010135	153433	A/T/G/C	[1, 5, 19, 20]
M9	AC003032	15424	C/G	[1, 5, 18-23]
M11	AC009977	456	G/A	[1, 5, 18-20]
M17	AC009977	2988	G del	[1, 5, 18-20, 22]
M45	AC010889	106136	G/A	[1, 5, 18, 20, 22]
M60	AC144429	75152	T/-	[1, 5, 18, 20]
M63	AC010889	108987	A/G	[1, 5, 20]
M69	AC010889	132407	T/C	[1, 5, 18, 20]
M74	AC010889	128116	G/A	[1, 5, 20]
M75	AC010889	128526	G/A	[1, 5, 18, 20]
M89	AC010889	155662	C/T	[1, 5, 18, 20, 22]
M112	AC010889	6308	G/A	[5, 20]
M122	AC010889	3023	T/C	[5, 18, 20-22]
M123	AC010889	2935	G/A	[5, 20]
M124	AC010889	2850	C/T	[5, 18, 20]
M132	AC010889	134610	G/T	[1, 5, 20]
M166	AC010889	3043	G/A	[5, 20]
M168	AC002531	5996	T/C	[1, 5, 18, 20]
M170	AC002531	39796	C/A	[1, 5, 18, 20]
M172	AC004474	7940	T/G	[1, 5, 18, 20]
M173	AC004474	64722	A/C	[1, 5, 18, 20]
M174	AC002531	146284	T/C	[1, 5, 18, 20]
M181	AC002531	43558	C/T	[1, 5]
M182	AC002531	61080	C/T	[5, 18]
M203	AC006376	148621	G/C	[1, 5]
M207	AC006376	139206	G/T	[1, 5, 18]

Name	Position (BLAST) Access #	BP #	SNP	References
M213	AC006376	83974	C/T	[1, 5]
M214	AC006376	29148	C/T	[1, 5, 18]
M215	AC006376	25047	A/G	[1, 5]
M216	AC010877	131816	C/T	[1, 5, 18]
M217	AC010877	131585	A/C	[1, 5, 18]
P25	AC006386	26234	A/C	[1, 18]
P29	AC002992	73192	G/T	[1, 18]
PN1	AC010137	5488	G/A	[5, 21]
PN2	AC010137	4152	G/A	[5, 21]
PN3	AC010137	3906	T/C	[5, 21]
RS035	AC024236	5093	A/G	[1]
RS041	AC007274	109042	G/C	[1]
RS155	AC011745	83585	T/C	[1]
RS182	AC007274	66178	T/C	[1]
RS394	AC013412	172634	C/T	[1]
RS407	AC010889	81439	C/G	[1]
RS422	AC006376	147897	T/A	[1]
RS640	AC006040	130257	G/A	[1]
RS653	AC006376	148760	G/C	[1]
RS664	AC006991	16948	TT del	[1]
RS734	AC009235	149562	G/A	[1]
RS763	AC010977	106894	C/G	[1]
RS843	AC007678	19581	A/C	[1]
RS917	AC012077	111712	C/T	[1]
RS924	AC006158	46998	A/C	[1]
SRY10831	AC006040	64927	A/G	[18, 21]
TSC389	AC012077	111712	C/T	[1]
TSC810	AC006040	76207	T/C	[1]

Loci “Names” are those assigned in reference materials or the last three numbers of their ID if taken from internet databases. “Position (BLAST)” represents NCBI nucleotide accession number and the base pair position within the accession for the SNP of interest. “SNP” represents the mutation observed or reported in references.

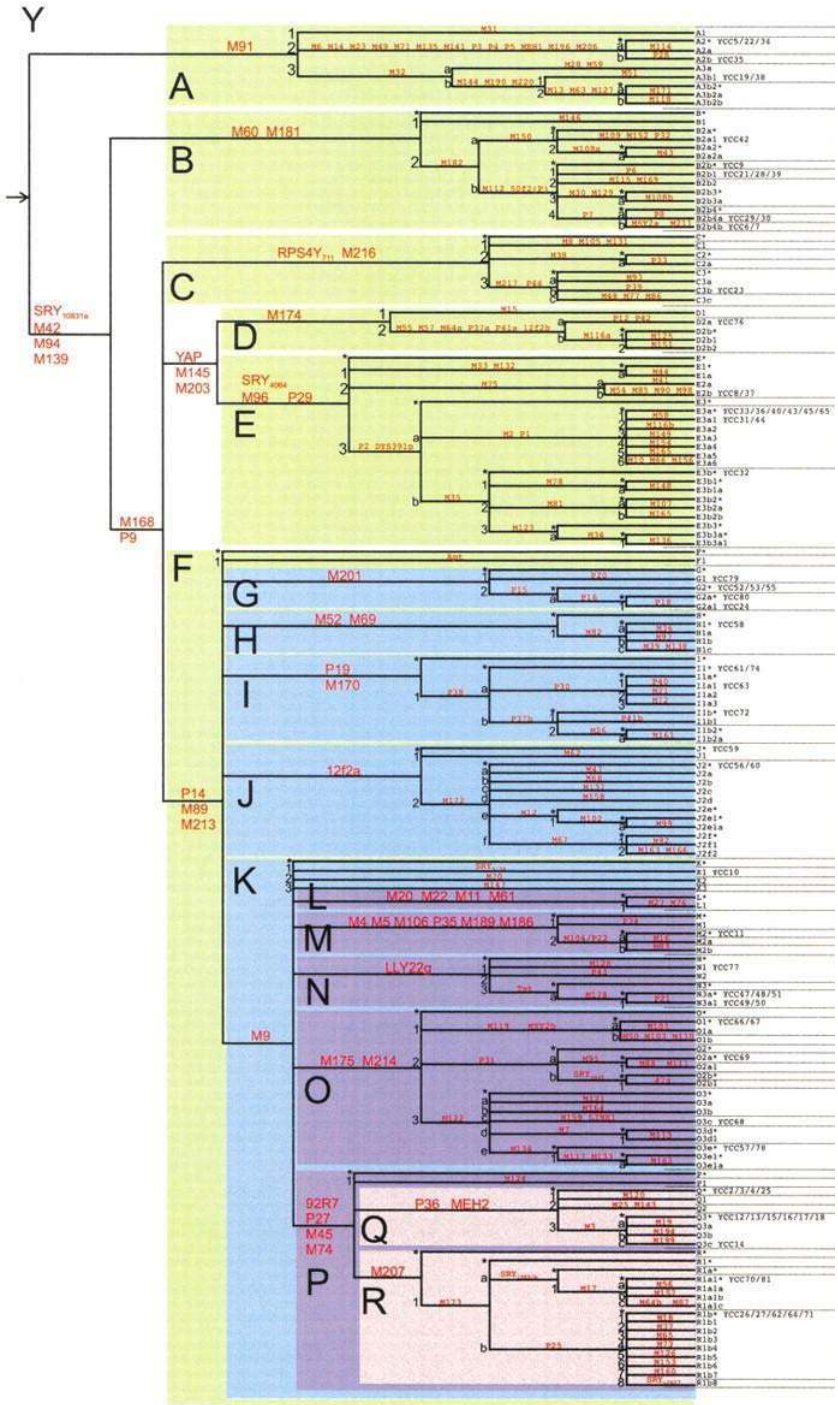


Figure 4. Y Halpotype Tree. Y Haplotype Tree designed by Underhill et al. [18]

Table 3

Loci PCR Primer List

Name	F Primer	Biotin	R Primer	PCR Product Size (bp)
M2	aagtccagacccaggaaggt	<	acagctccccctttatcctc	162
M3	aaacttcagacctccagggacattaa	<	ctgctgccagggctttcaaata	303
M6	tgcttcccctactttgatgg	<	tctgggagttgggagttagg	179
M9	gcgctaccttacttacataac	<	gtttggtgcctaaagattaga	190
M11	cacgagcttcaaaaatgagga	>	ctgggggaagcaaatactga	352
M17	gagtttgtggttgctggttg	>	cttgtttctccaggcttttca	167
M45	ggggtgtggactttacgaac	<	acagtggcaccaaaggtcatt	179
M60	gccctgatgtggactcaacc	<	gaggagcgacgtctacgaaag	235
M63	ggctcagaccaaaaattcca	>	aggtgggaagttggaggatt	175
M69	gggtagcctgttcaaatca	>	tggcatgaagcatgtaagga	244
M74	ggaaagtctgaaaaataatca	<	tcctgttagctagaaccttac	224
M75	acttgtcaaaagccaaaaca	>	agatcgcaccattgcagtc	199
M89	gcccacagaaggatgctgct	<	aggctggcacactttgggtc	134
M112	atggaactattttattacacc	>	aaagcaaaagagaactgcctc	192
M122	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	393
M123	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	393
M124	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	393
M132	ccaaaagttgaatggtaaaacaga	>	cacccatacacaaaaatatctgga	200
M166	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	393
M168	caagtaccagattaaggcact	<	ccccagcttagtaattaactc	333
M170	tgggttccagattttgcat	>	tgagacacaacccacactgaa	232
M172	tgcctctcagtatcaacaggtaaa	>	tcactccatgttggtttgga	179
M173	aaagttgatgccacttttcaga	<	gcagttttcccagatcctga	242
M174	tctccgtcacagcaaaaatg	>	cctggagatgcaaaaggaga	192
M181	ggcttgggatttttctcctg	>	tggaagaaacaacaatgacca	217
M182	gcccctgtgtgtatgtgatg	>	tgtgccatcacaccatcttt	207
M203	agagtatcgccagccaaccag	<	aagaaaatggcggaagg	137
M207	gacggcatatttacatga	<	ggaaggacatgccttttct	347
M213	ccaaaccaaggccatataaaaa	>	aaatattcagaacttaaaacatctcgt	139

Name	F Primer	Biotin	R Primer	PCR Product Size (bp)
M214	taggctgattttgctgctga	>	gaggtgggagggttgctt	195
M215	catcccatgaaatatacacagaaa	<	ggaacaactgcgagcaaata	214
M216	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	339
M217	tcctgcagcgaattagatt	>	tcctgcagcgaattagatt	339
P25	gtgaattatctgcctgaaacctg	<	gatggaccgagatacgagaca	60
P29	gcactccaatctgctgacttaaa	>	aaataccagtcgtcaccctctct	196
PN1	tcccctgctattgctagtcc	>	agggtcttgagagggagagc	180
PN2	agctccagccatcttttcct	>	tgagtcccttgatgcaaatg	209
PN3	cccaatccaatgctcctaac	>	taccaagaggcaggggtaca	112
RS035	gtgaagggtggatgggttc	<	ggccatagtgtggactgctt	122
RS041	g ttctcacttt gccagcttca	<	gcccaaggttctatctccagt	162
RS155	gttcctgcatttttgggattt	<	ccaaaggcaatatcttcagca	231
RS182	tgtgagtgtaggcatgttcca	<	tatttccctccggttgtcatt	322
RS394	caccattttcacaggatttgc	<	caacaccatcagccaaacctc	282
RS407	ccaggtgattccctatggtg	<	acttcttcgcatgcagctct	175
RS422	aggctaaacatacacgcatca	>	ggctaaaccgggtctacacca	282
RS640	aggtttcccaaatgtctaaat	<	ggtagcggatggttcgagcat	288
RS653	aacagcggcggtagtgag	<	gcagcgattgaaggcgtcttt	116
RS664	ggtgtgtttgctttcttctcc	>	gtcttgctccttgccagttc	217
RS734	ttggattgggtttagaagatgg	>	caaatccaactaatgccacca	157
RS763	tcaatgttgcctacgtcctg	>	gccaatagcatatagcatagc	565
RS843	tccctcatgacaccatcaaa	>	aatgaaaatagctgccaagtaaaa	74
RS917	gtacataccagtgatcattcacc	<	agaaagtttgagctttagactcc	321
RS924	aggacggcagtactcagcat	<	tgacacccgaataaagtgga	225
SRY10831	agtggaaatatagttggctca	>	tcattcagtatctggcctctt	244
TSC389	agaaagtttgagctttagactcc	>	gtacataccagtgatcattcacc	330
TSC810	tttaacggttttaaactaagatcc	<	ctatcggtaattcaacaaaac	262

"F Primer" represents the nucleotide sequence for the forward primer. "R primer" represents the nucleotide sequence for the reverse primer. "Biotin" points to the primer which has the Biotin label used for pyrosequencing: "<" indicated the forward primer carries the Biotin label, ">" indicates that the reverse primer carries the Biotin label. "PCR Product Size" is the size in base pairs of the PCR product using the listed primers.

Table 4

Extension Primer List

Primer Name	Sequence	Sequence to Analyze
92R7E (’02)	tgaacacaaaagacgtagaa	g/a tttgtctttgctggtcatatttaac
92R7E ('01)	atatgaccagcaaagacaaa	c/t ttctacgtcttttgtgttcatg
M112E(2)	aggatgcagtactgataaaa	a c/t actgctttgtttttat
M11E	ctccctctctccttgtattc	taac a/g gaaaggtttagaacttgcataa
M122E(2)	attttcccctgag	agc a/g tgaattagtatctcaattgct
M123E(2)	tctaggtattcaggcga	tg c/t tgatatgctagttcagat
M124E(2)	gcacaaactcagtattatta	aacc g/a acttccctgttccccccata
M132E	aacacaagcgtgagcatt	aata t/g aaccatataaatgcatttttttgaa
M132E(2)	cacaagcgtgagcattaata	t/g aaccatataaatgcatttttttgaa
M137E(2)	catattgagattttgctttc	ct a/g taaagcaaagttgag
M166E(2)	agcgtgaattagtatctcaa	ttg c/t ttttctaccttgat
M168E	ttgttttaattcttcagcta	gc g/a caactcacctccaacacatactccac
M170E	ttacttaaaaatcattgttc	a/c ttttttttcagtgtgggttgtgtc
M170E(2)	tatttacttaaaaatcattg	ttc a/c ttttttttcagtgtgggttgtg
M172E	acccattttgatgctt	t/g acttaaaaggtcttcaattatt
M172E(2)	tcccccaaacccattttgat	gctt t/g acttaaaaggtcttcaattatt
M174E	taccttctggagtgccc	t/c agtgcagaagtgaggggtgc
M17E	aaccccaaaattcacttaaa	aaa [c] cccgtaacaaccagcaacca
M17E(2)	tggttgctggttgttacggg	[g] ttttttttaagtgaattttgggg
M181E	caacttgatcatctttttg	a c/t tgctttaaggtagtagcttg
M182E	gcagtggttaatgtaaacaa	a c/t gtaataaattatgtggtatt
M2E	ttcattgttaacaaaagtcc	a/g tgagatctgtggaggataa
M2.2E	tttatcctccacagatctca	t/c ggacttttgttaacaatga
M2.2E(2)	ctttatcctccacagatctc	a t/c ggacttttgttaacaatga
M203E	gggaggcgcttggtggca	tg [g] acaggtacgcacctccatca
M207E	caaaaggtattgttattctc	ttt [c] taaatttcttgcttgacttacattt
M207E(2)	aaggtattgttattctcttt	[c] taaatttcttgcttgacttacattt
M213E	attacatatattaataagaagtca	c/t gtaacgagatgttttaagttctgaat
M213E(2)	attacatatattaataagaagtc	a c/t gtaacgagatgttttaagttctgaat
M214E	ctttcgttcgtttatttttc	t/c gttgttttcagacagtgtctca
M215E	cagctggaacagttagaaag	t/c cagtttgtcttaatgcagcaagtat

Primer Name	Sequence	Sequence to Analyze
M216E	agttatgtatacctgttgaa	t g/a ttacatttctttaataaaggaattt
M216E(3)	gctagttatgtatacctgttga	at g/a ttacatttctttaataaaggaattt
M217E	atttttccttctgaagagtt	t/g gtgtcacccaacttttttcattctc
M3E	tgggtcacctctgggac	tga c/t aattaggaagagctggtacctaa
M3E(3)	atgggtcacctctgggactg	a c/t aattaggaagagctggtacctaa
M3E ('01)	ggtaccagctcttcctaatt	gtc a/g gtcccagaggagaccca
M42E	tcagatttaggacacaaaag	c a/t actacataatgaaaaagagagc
M45E	gcagtgaaaaattatagata	a/g gcaaaaagctccttctgagg
M45-2E	cctcagaaggagctttttgc	t/c tatctataatttttcactgccaatt
M60E	agtaaccactgtgtgcc	tgat [a] cttaagtcatgcattttgaaatgt
M63E	gacacgctcaggtacctcaa	a/g gaatcctccaacttcccaccttcac
M69E	ggctgtttacactcctg	aaa c/t aaaatatatttcagcaag
M6E	gttggcttgtagttctttct	a/g/c/t ggaaaaatattattctaatttcc
M74-2E	cttaaagcaacttaaaaatg	t/c accagtttttaaaataccta
M74E	taggtattttaaaaactggt	a/g cattttaagttgctttaag
M74E(2)	tacttaaagcaacttaaaaa	tg t/c accagtttttaaaatacc
M75E	acaattatcaaaccacatcc	g/a tatatatacagagaaat
M75E(2)	aaagacaattatcaaaccac	atcc g/a tatatatacagagaaat
M89-2E	actcaggcaaagtgagagat	a/g agatttttgtacataacctta
M89E	aaggttatgtacaaaaatct	c/t atctctcactttgcctgagt
M9E	attaaaagaaaaataaagag	c/g attcaaccatcttaggccgt
M9.2E	acggcctaagatggttgaat	g/c ctctttatttttctttttaat
M9E(2)	ttgccggattctaccaactta	c/g gagaaataaaaagaaaatta
M9E(3)	aacggcctaagatggttgaa	t g/c ctctttatttttctttttaat
M9E(4)	aagaaacggcctaagatggt	tgaat g/c ctctttatttttctttttaat
P25E	tacgagacacaattctattt	t/g caggcaggtttcaggcagataattca
P29E	caatctcatctaaaaatacctt	ta c/a agcaacatctagtgtttg
PN1E	ttccccttttaagacaaaa	g/a gctctccctctcaagacccttg
PN2E	gaggtgcccctaggaggaga	a a/g gcattagttttattaatgctccttt
PN3E	taactctaatttacctcccc	c/t tcaagcagaggcagttt
RS035E	tctttttgcagttgaagttg	ca a/g gaacccatccacccttcac
RS035E(2)	tgcagttgaagttgca	a/g gaacccatccacccttcac
RS035E(3)	atcttttgcagttgaagtt	gca a/g gaacccatccacccttcac

Primer Name	Sequence	Sequence to Analyze
RS155E	atatcttcagcacaaaagcc	taa t/c agagaaaaacttcaaacctaaacc
RS182.2E	tcctccccaaaccaaggta	caca a/g aggtacacaggcacaggtg
RS394.2E	ctgcctccaatgttcccc	ctgc a/g gcccctgccccccatacttcctgat
RS407E	caatatgcacccttgctaaa	c/g gaccaaatgagacactgc
RS407E(2)	ttggcaatatgcacccttgc	taaa c/g gaccaaatgagacactgc
RS422E	cagtttataggtcaaatatc	tac t/a gcaaactcttcaccgctgtaact
RS624.2E	aatattaacagatgacaaag	t/g gttctaaatgcccttgaattg
RS636E	aggtgaaccttgaaaatgtt	a c/t actgtgtgaaaaagtcagatacaa
RS640.2E	ggagtgctcttgggaaaaa	cta t/c attaaccagtcacagagctgatgatc
RS653.2E	gtttccatgaaatc	c/g tgcgcagtgtcgctcactaccgccgctgt
RS664E	tcagatgacaaagagaaact	ttactatat [tt] gggggagtaatgtttgatatt
RS664E(2)	tgacaaagagaaactttactatat	[tt] gggggagtaatgtttgatatt
RS664E(3)	acaaagagaaactttactatatt	[t] gggggagtaatgtttgatatt
RS678E	acatgccttctcacttct	c [ttctc] agaatgaacagaaacaaaggtat
RS734E	ctgcagggcaccaccagc	c g/a ccttggtggtggcattagttggatttgg
RS843E	atgcactgttgtaaagcc	tg a/c gtattttacttggcagcta
RS924.2E	ccgaataaagtggataaaca	t/g acacacacacacacacacacacaca
SRY10831E	aggtgaaccttgaaaatgtt	a c/t actgtgtgaaaaagtcagataca
SRY10831E(2)	ataggtgaaccttgaaaatg	tta c/t actgtgtgaaaaagtcagataca
Sy81E	ctttatcctccacagatctc	a t/c ggacttttgttaacaatga
Sy81E(2)	tttatcctccacagatctca	t/c ggacttttgttaacaatga
TSC389E	ctctgcaactaaaaatgaag	t/c acatcctgggtgaatgatcactggtat
TSC810.2E	caactgaaaataccagtatc	a/g caacttttttgttcaaatcatac
TSC810E	atgatttgaacaaaaagttg	t/c gatactggtattttcagaag
TSC810E(2)	gccaactgaaaataccagta	tc a/g caacttttttgttcaaatcat

"Primer Name" represents the name given to the extension primer for identification. Sequence represents the sequence of the listed primer. Sequence to Analyze" represents the nucleotide sequence that would be dispensed and examined during pyrosequencing (excluding blank nucleotide dispensations).

Table 5

Outer PCR Primer List for 13plex Primer Set

Name	Outer F Primer	Outer R Primer	PCR Product Size (bp)
M3	aggaatcacttggctgtcactt	taatcagtctcctcccagcaag	485
M9	tctgaccctaagcaaatccaa	gacggcgtttaaaacaatcg	443
M74	tgagcagaaaaaggttttcaat	gcaagttttcttgatgtgtgga	395
M170	acagggttggctagtggatct	acgaaggacacaaaacctctg	383
M172	atgagccctctccatcagaa	gcaaatagcaaaaggggggta	376
M207	aagggcaagcaaaatagcaa	cgtgcagatgtatgtattcttcc	434
M214	tggcatcttgaggttgattg	tgatctcttgagtctggggttt	434
P25	ccaaagcctagaatgaaattgtg	caagacaaaggctaaagcaaaaa	369
PN1	ggccaactccaaccaagtta	ccatctgttgatgaggtcca	339
PN2	ctcacattgtcccacctcct	ggctgcatcatccatagctt	404
RS640	ttcagctgagcactcttcca	cgttcaccttgatgacagga	372
RS843	ttactgccgaaagcttgaca	ccaaccatttccatgtagcc	427
TSC810	ggctttacgaggattagacagg	accccgctaatgctaccttc	409

Loci "Names" are those assigned in reference materials or by present lab if taken from internet databases. "Outer F Primer" represents the nucleotide sequence for the forward primer. "Outer R Primer" represents the nucleotide sequence for the reverse primer. "PCR Product Size" is the size in base pairs of the PCR product using the listed primers.

3.2 SNaPshot Analysis

3.2.1 General

Y chromosome SNPs were analyzed using either SNaPshot or Pyrosequencing. SNaPshot analysis was performed with a single source male sample and loci analyzed in a single tube and therefore produced very straight forward results with a single peak at approximately 2900-3200 data points on the time scale. A positive result was arbitrarily defined as a single

peak height over 200 RFU. Figure 5 shows an example of a male sample (ME) when amplified showed only one amplification product. Therefore when analyzed using SNaPshot produced only one detectable nucleotide extension. In this case the male individual (ME) would be classified as having the A allele at the M2 locus using the M2E extension primer.

Although the allele calls using SnaPshot are very straight forward, they provide very little assurance as to the location of the extended nucleotide. When a peak is observed due to the extension of the extension primer, it is unknown whether the extended nucleotide was added with proper extension primer binding or to the extension primer bind to locations other than expected.

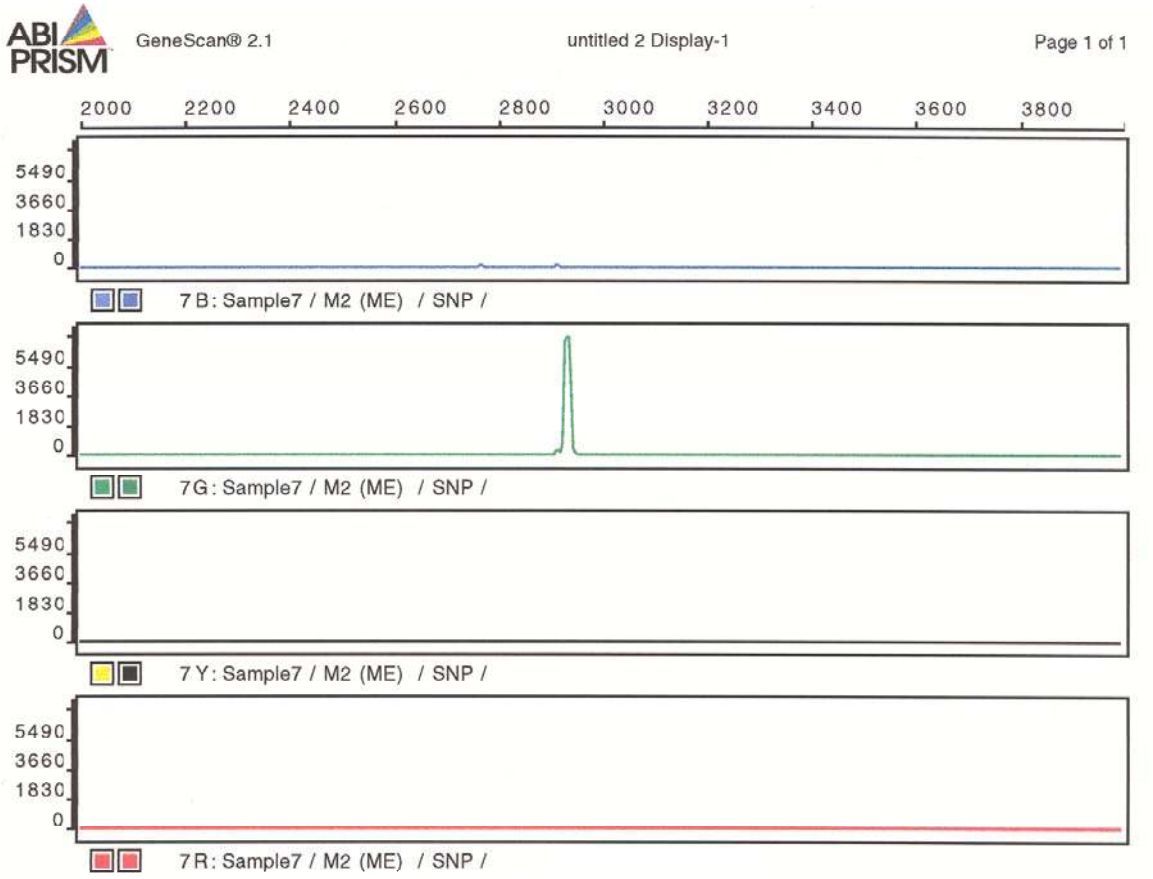


Figure 5. Standard SNaPshot Reaction. SNaPshot analysis using ABI 310 Genetic analyzer of male individual (ME) at the M2 locus using M2E extension primer showing only single extension of the A (green) nucleotide.

3.2.2 Extension Primer Choice

With single base extension, there are only two options for the extension primers since the SNP must be adjacent to the 3' end of the primer. In most cases, either one of the primers will work adequately for SNP analysis. Unfortunately, there will be instances that one or both of the possible primers are not suitable for SNaPshot single base extension due to either mispriming or lack of sufficient extension product. When examining the M9 locus, it was observed that the

M9E extension primer showed very minor or no extension products (Figure 6) with multiple individuals. When examined using the same amplified product and the reverse primer M9.2E an extension product was always observed, but it was sometimes below the 200 RFU threshold (Figure 7). SNaPshot analysis was eliminated in favor of Pyrosequencing, to overcome the lack of flexibility found in extension primer selection and provide information about what nucleotide is being examined.

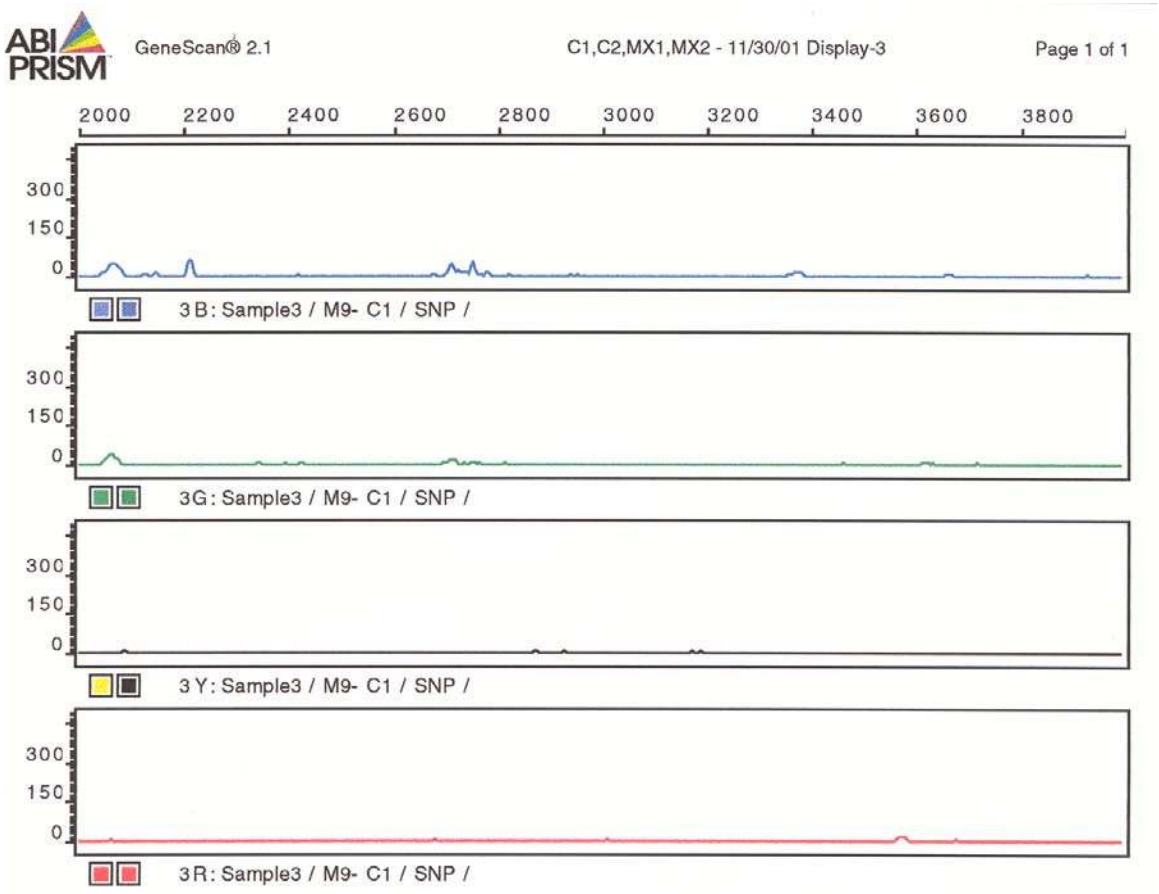


Figure 6. M9E Extension. SNaPshot analysis using ABI 310 Genetic analyzer of male individual (C1) at the M9 locus using M9E extension primer showing no nucleotide extension.

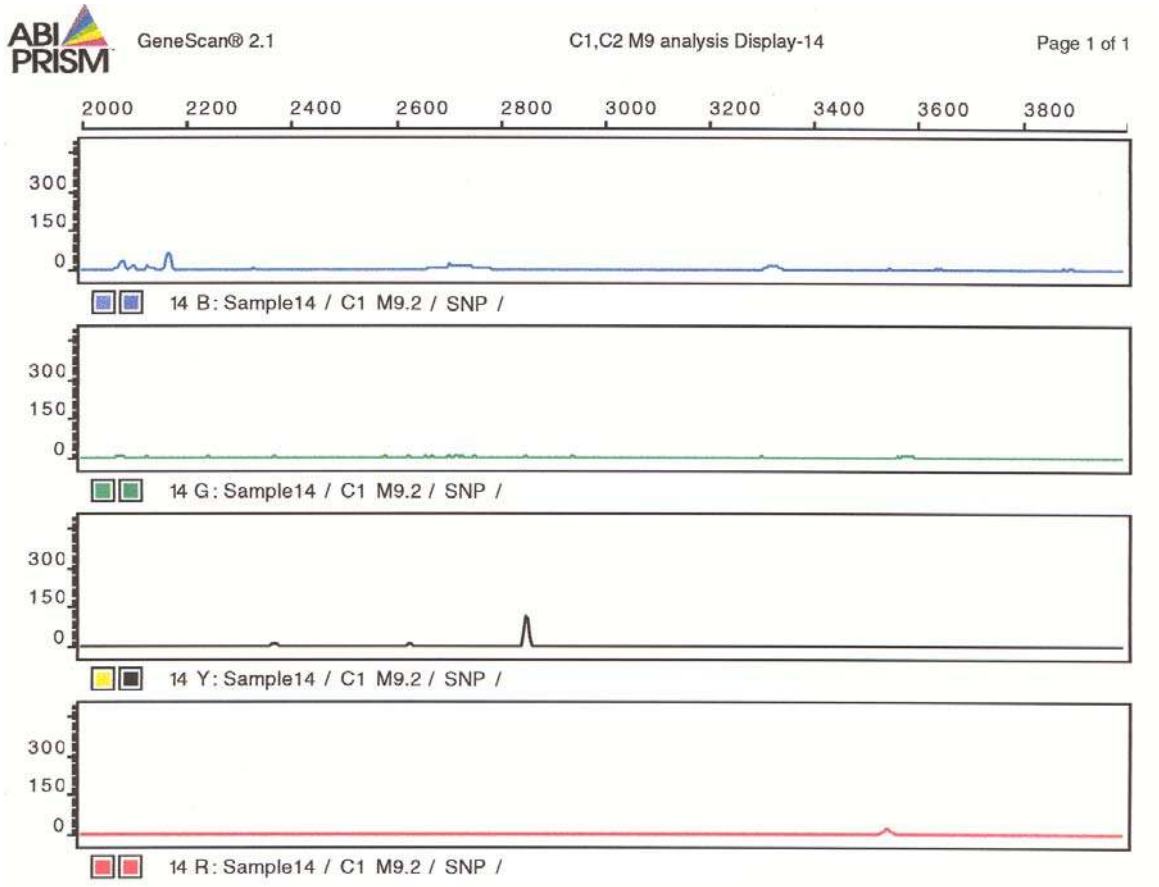


Figure 7. M9.2E Extension. SNaPshot analysis using ABI 310 Genetic analyzer of male individual (C1) at the M9 locus using M9.2E extension primer showing nucleotide extension of C (black) nucleotide below 200 RFU threshold.

3.3 Amplification

After the designing of the initial PCR primer set, primers were amplified using male, female, and no DNA. This method ensured that male amplification produced only one distinct PCR product of appropriate size, and the use of female or absence of DNA did not cause amplification. PCR primer sets that caused multiple male products, no male products, or

products in the female or blank lanes were redesigned. Loci that continued to show multiple male products, no male products, or products in the female or blank lanes were eliminated from the development process. The following loci were removed due to female amplification: RS394, RS678, and RS924 (Figure 8), RS182, RS041, RS917, RS763, and RS743 (data not shown). All other loci which showed amplification of only one male product (at the appropriate size), and no female or blank DNA amplification products were eliminated from further evaluation. Figure 9 shows examples using M170, M216/7, and RS664 of what the proper amplification pattern should be with male amplification of only a single locus.

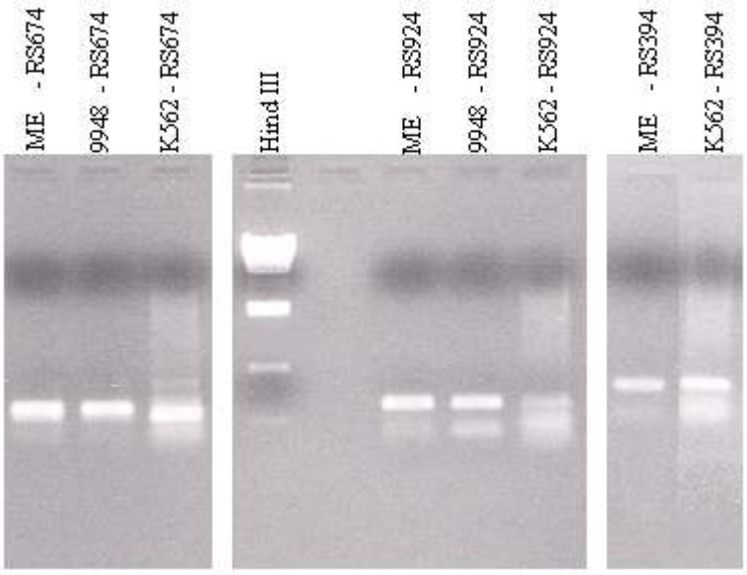


Figure 8. Female Amplification Gel. 1% Agarose gel containing 0.01% EtBr run at 200V for 15 min. Electrophoresis: run post amplification of male (ME, 9948) and female (K562) samples using RS674, RS924, and RS394 PCR primers.

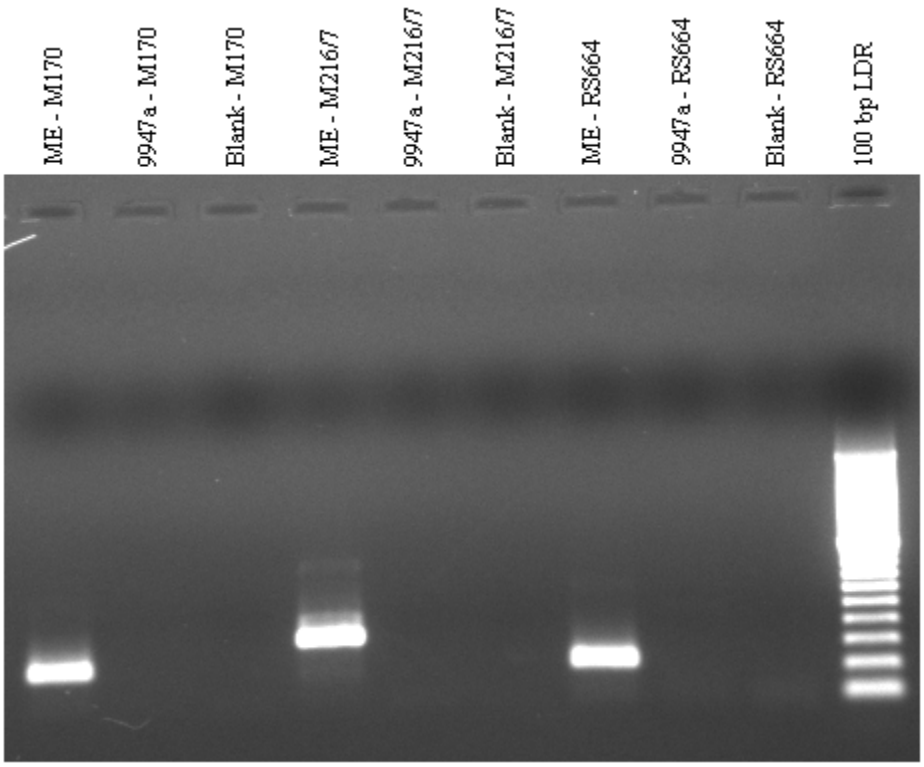


Figure 9. Male Amplification Gel. 1% Agarose gel containing 0.01% EtBr run at 200V for 12 min. Electrophoresis: run post amplification of male (ME), female (9947a), and no DNA (Blank) samples using M170, M216/7 and RS664 PCR primers.

3.4 Pyrosequencing Analysis

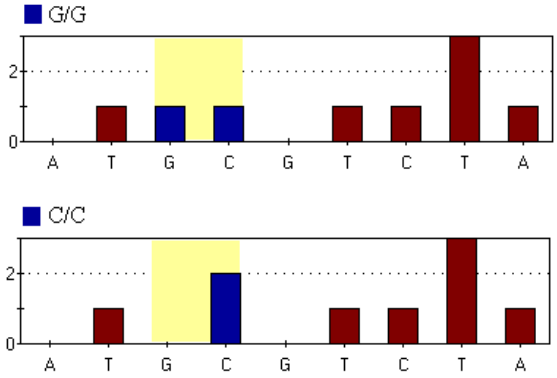
3.4.1 Singleplex Analysis

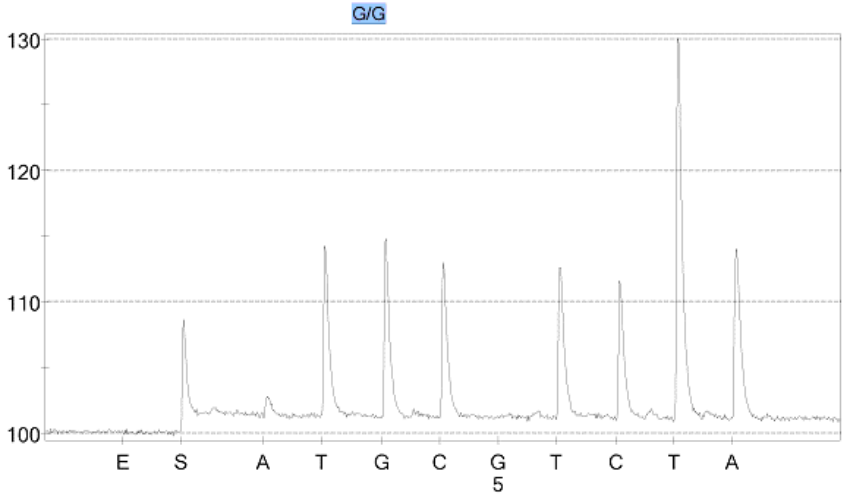
For singleplex (one SNP) pyrosequencing reactions, there are three possible outcomes the program recognizes, but because the Y chromosome is haploid the heterozygote outcome is eliminated in cases of single source analysis. Therefore, the sequenced pyrogram is compared with two possible outcomes for SNP determination. If the pyrogram matches exactly to one of the possible outcomes, the Pyrosequencing 96MA SNP Analysis Software will identify the SNP

and mark the well as passed (blue) (Figure 10). If the produced pyrogram does not match exactly to one of the possible outcomes due to: peaks appearing in negative control blank nucleotides dispensations (Figure 11), ambiguous allele calling peaks (Figure 12), non reference peak heights (Figure 13), too large a substrate peak (Figure 14), or dispensation errors (peaks appearing between dispensations of nucleotide) (Figure 15) then the Pyrosequencing 96MA SNP Analysis Software will identify which SNP it believes the pyrogram best fits and marks the well as either to be hand checked (yellow) or failed (red) due to the severity of the inconsistencies with the expected results. These checked or failed reactions were routinely corrected with little effort. Peaks appearing in negative control blank dispensations were usually cased by template loops, primer dimmer formation, or alternate binding sites for the extension primer that were corrected with primer redesign. Ambiguous allele calls were usually the result of female amplification and therefore eliminated. Non reference peak heights were usually caused by steric hindrance if the SNP was located to close to the ends of the PCR product and were eliminated using primer redesign. Substrate peaks that were too high were eliminated by increasing the amount of washing buffer used in sample preparation. Dispensation errors were caused by clogged reagent cartridges that would dispense nucleotides into incorrect wells at incorrect times and was eliminated by changing to a new cartridge. If a locus continuously gave a hand check or failed analysis, after all efforts were made to eliminate the cause, then the locus was eliminated from consideration.

Theoretical outcomes for simplex entry M9

Sequence to analyze:
TC/GCTCTTTATTTTTCTTTTAA





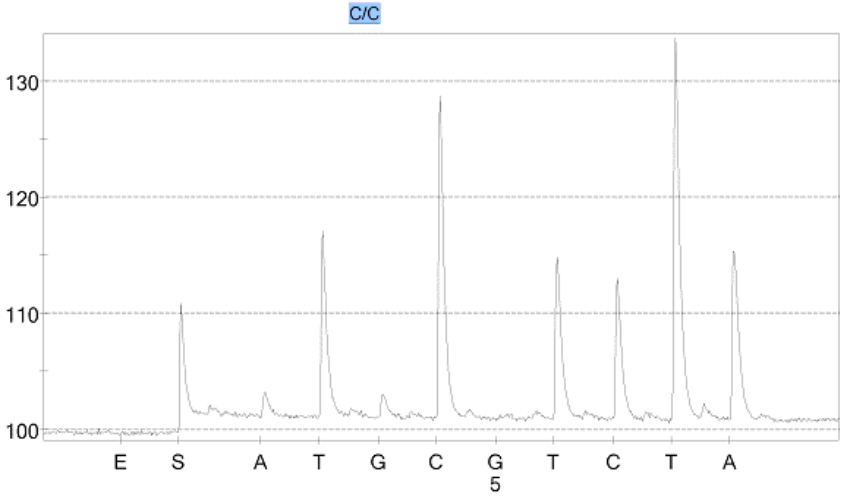


Figure 10. Passed-General Pyrograms. Two different pyrograms for two different male individuals (C = ME, G = 9948) amplified at M9 locus and pyrosequenced with the M9E(3) extension primer. The C allele will appear as a double peak at the C dispensation (#4) and a blank peak at the G dispensation (#3), the G allele will appear as a single peak at both the G dispensation (#3) and the C dispensation (#4).

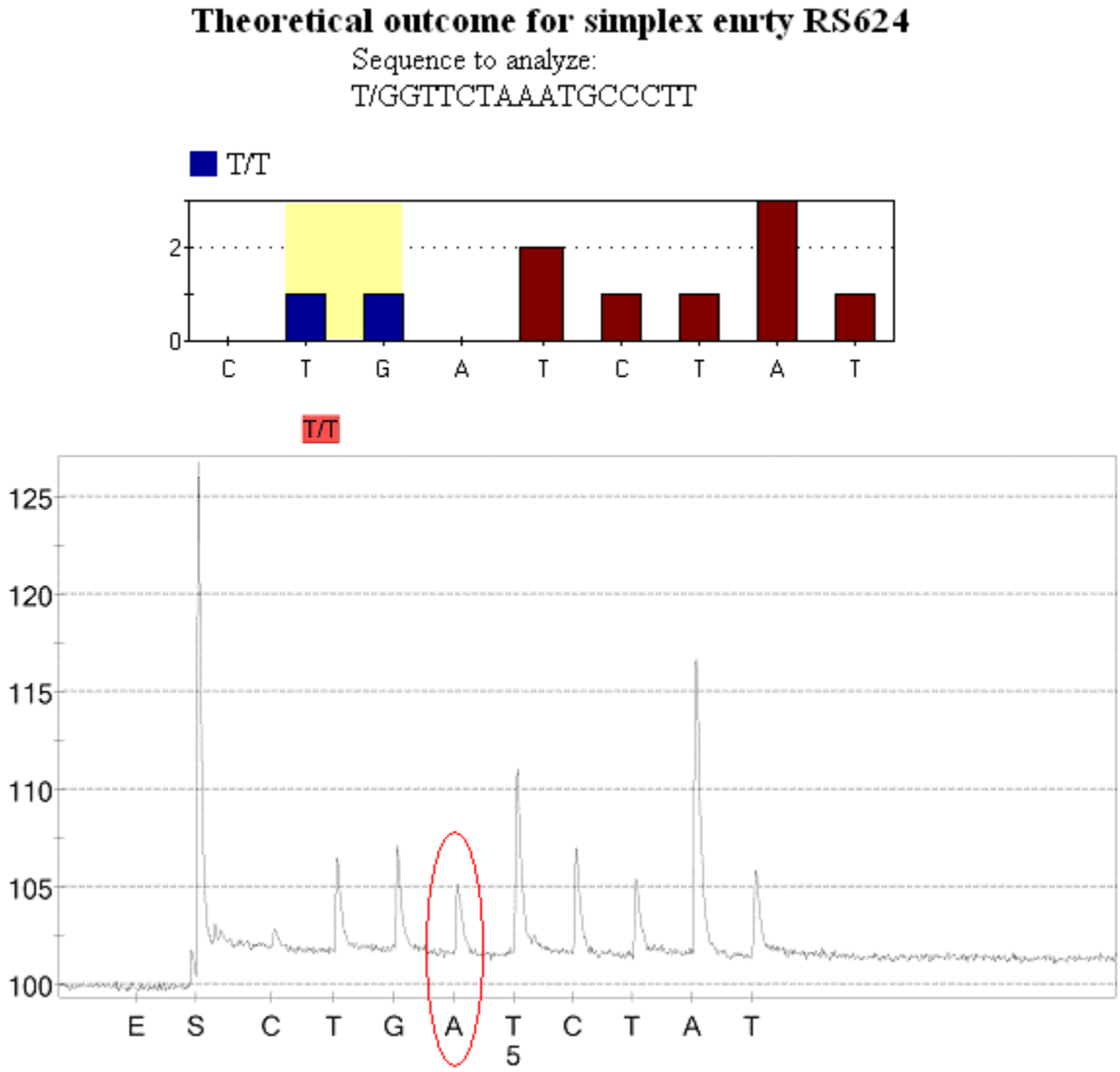


Figure 11. Failed-Peak in Blank. Male sample (ME) amplified at RS624 and pyrosequenced using RS624.2E as extension primer. Failed due to expected blank dispensation at A nucleotide (#4) shows a peak height of approximately 3-4 RFU.

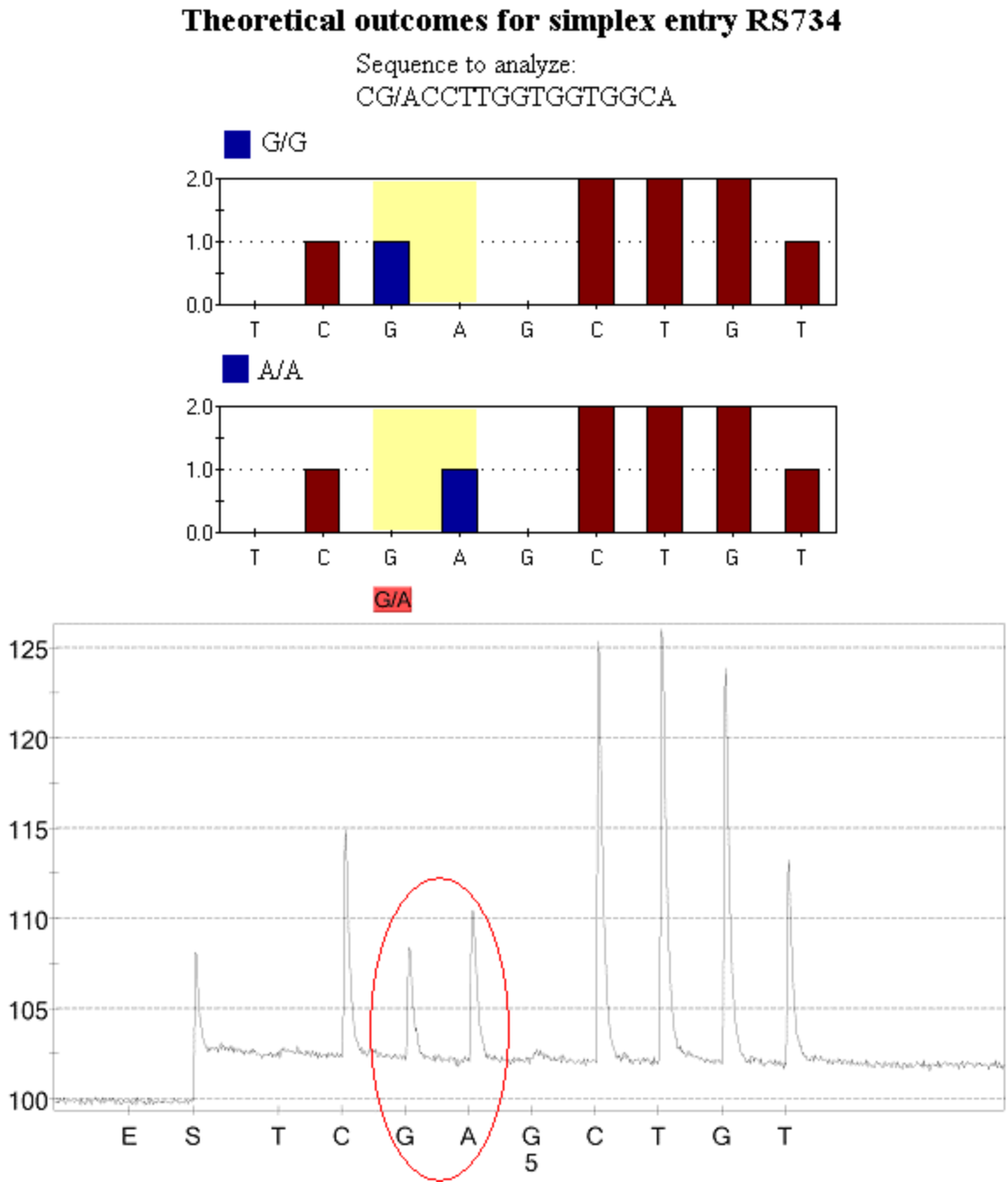


Figure 12. Failed-Ambiguous Allele Call. Male sample (ME) amplified at RS743 locus and pyrosequenced with RS743E extension primer. Ambiguous allele call due to peak in both the G nucleotide (#3) and the A nucleotide (#4), when theoretical outcomes predict one or the other.

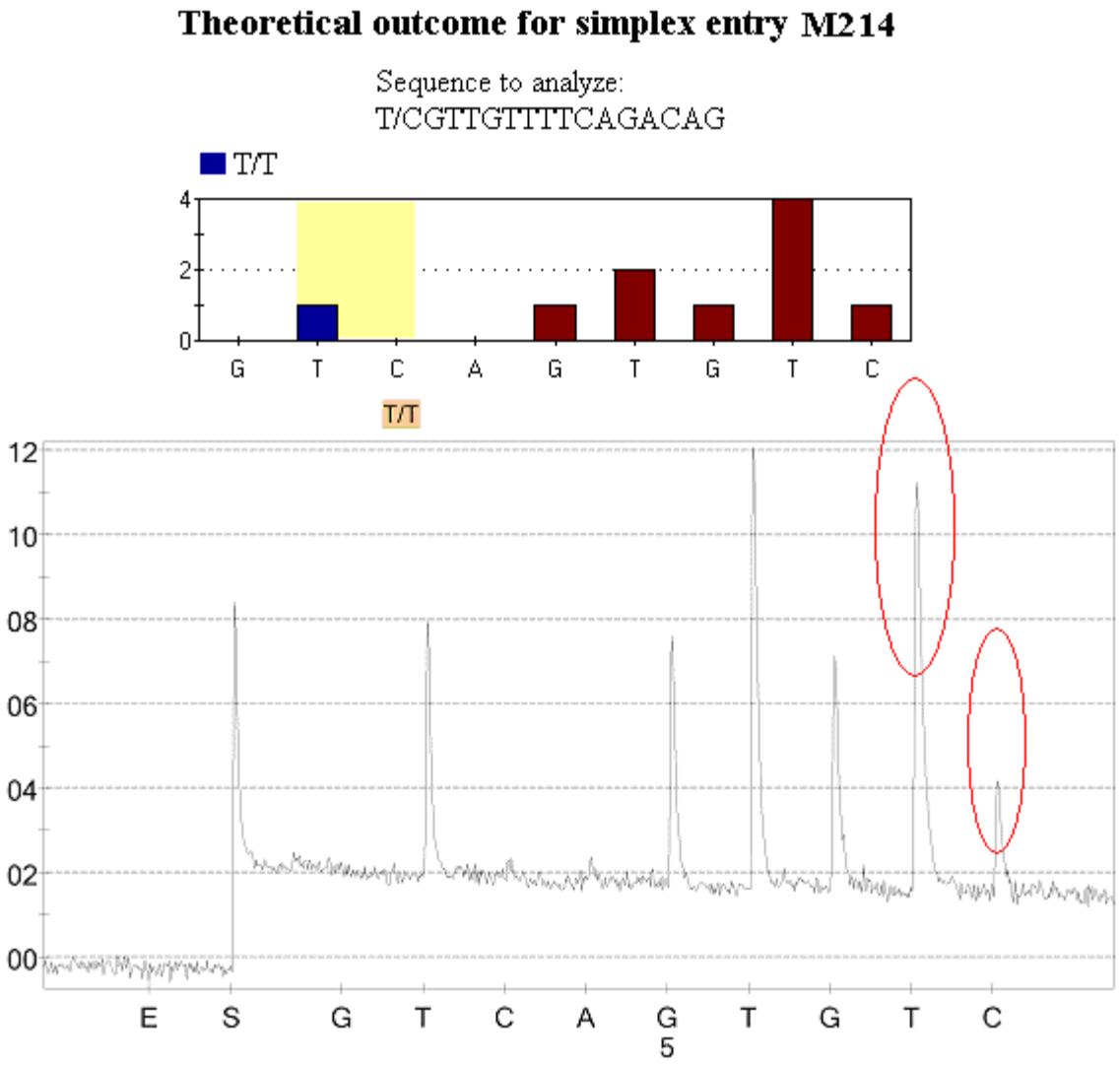


Figure 13. Failed-Non Reference Pattern. Male sample (C1) amplified at the M214 locus and pyrosequenced using the M214E extension primer. Pattern does not match reference at T nucleotide (#8) which should be approximately twice as high as T nucleotide (#6), and at C nucleotide (#9) which should be approximately the same height as G nucleotide (#7).

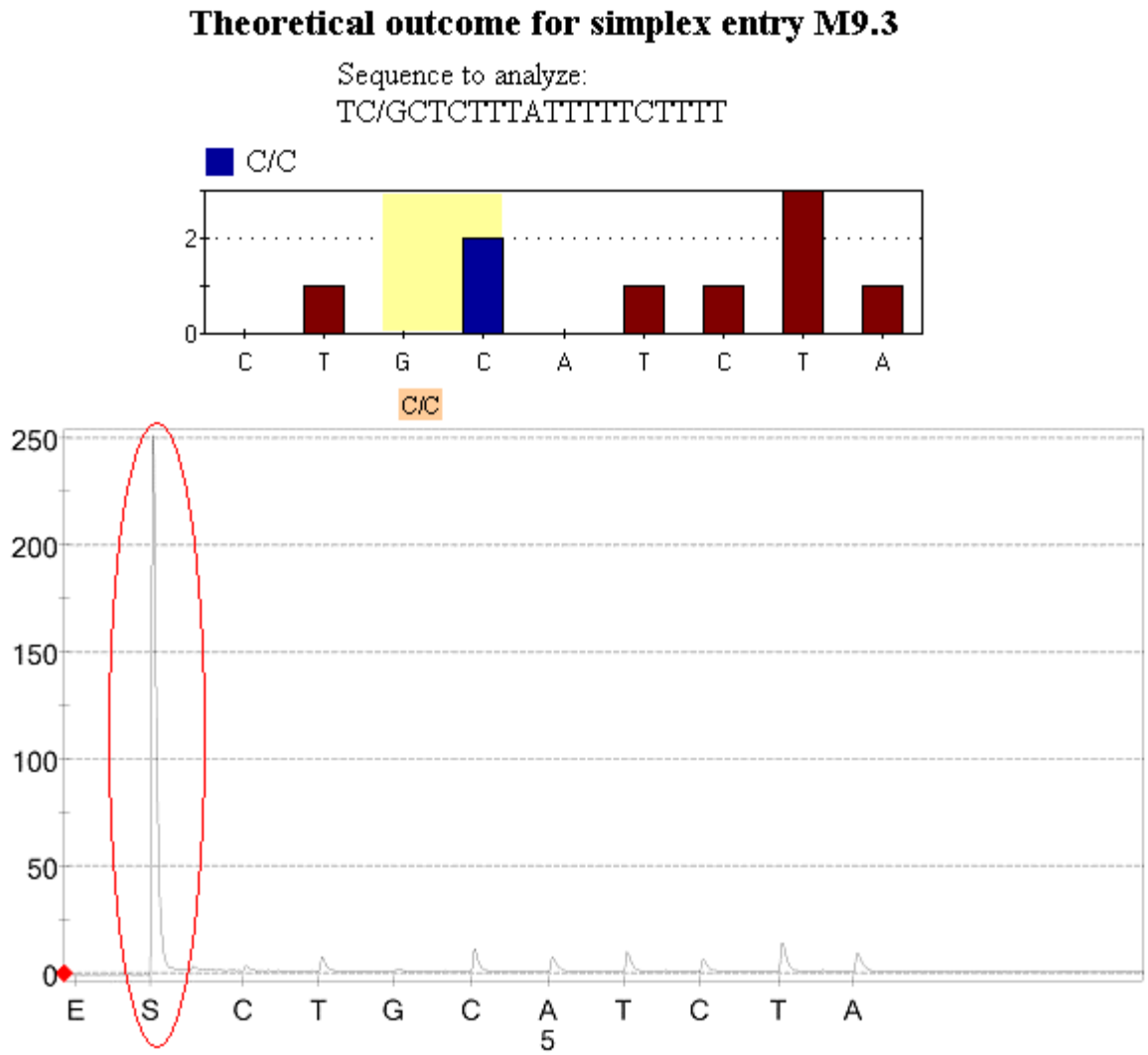


Figure 14. Failed-Substrate Peak Too High. Male Sample (C1) amplified at M9 locus and pyrosequenced using M9E (3) extension primer. Substrate peak is so high that the rest of the pattern is lost.

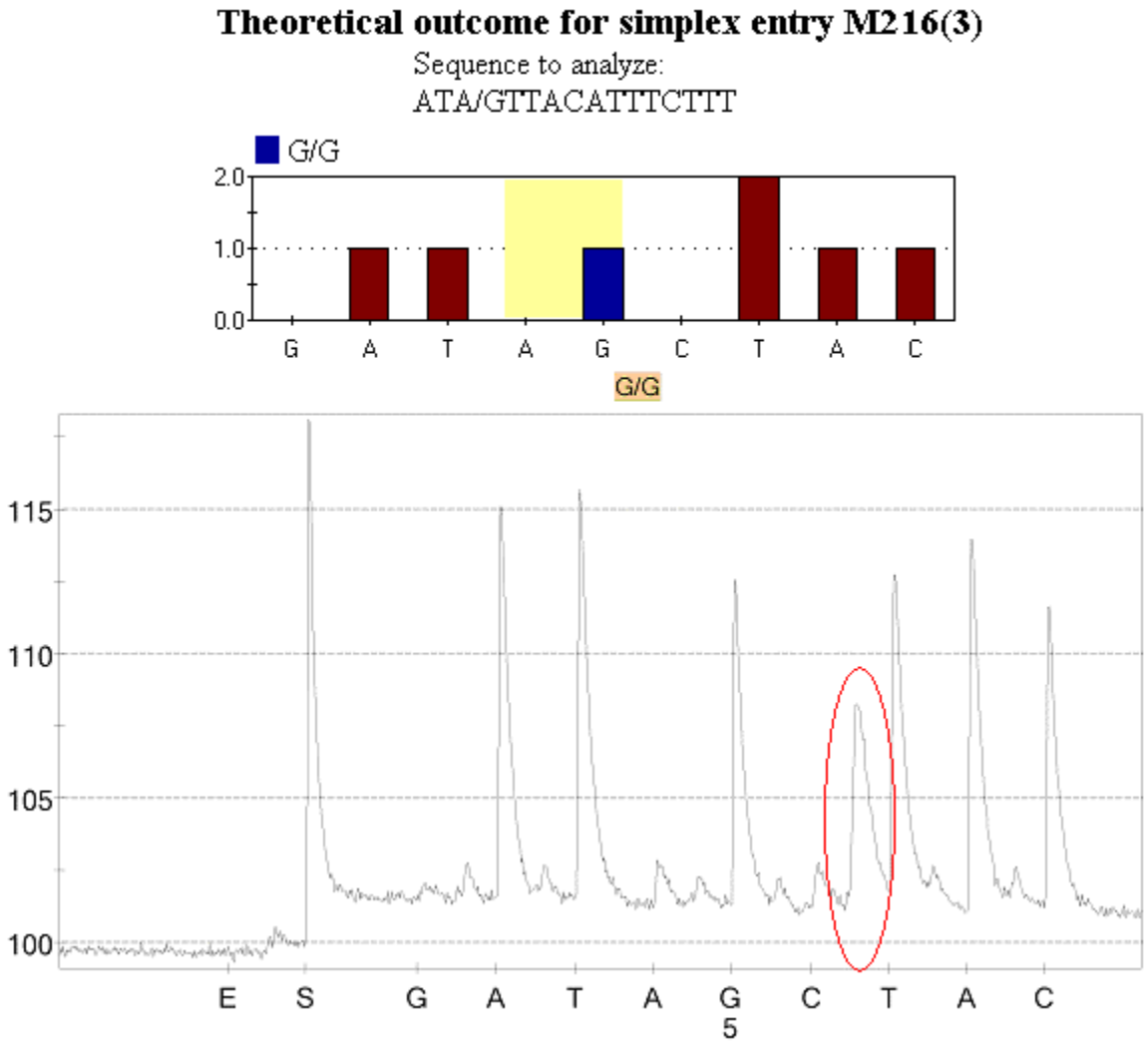


Figure 15. Failed-Dispensation Error. Male Sample (576) amplified at M216/7 locus and pyrosequenced with M216E(3) extension primer. Dispensation error occurred when some T nucleotide was inadvertently added to this reaction well (most likely when dispensing adjacent wells) between the C nucleotide (#6) and the T nucleotide (#7) causing a peak when no nucleotide should have been dispensed.

3.4.2 Multiplex Analysis

Pyrosequencing Multiplexes analyze two or three SNPs within the same reaction vessel in a single nucleotide disposition pattern. The original intention for the Y chromosome SNP muliplex analysis system was to amplify up to 50 SNP in one large volume PCR reaction using biotinylated primers, then using multiple aliquots of that PCR product for pyrosequencing. Amplifying more than three loci in a single reaction lead to problems with high background peaks, nonreference sequences, and peak height balancing due to template-template interactions. Figure 16 shows an attempted four loci amplification (M17/M3/RS640/ M216/7) in which the templates alone, when mixed in a single reaction vessel, produced a pyrogram due to binding between the multiple templates. Almost all amplification products of more than three loci caused significant pyrograms (peak heights greater than 1.5), when only the PCR products were examined (in the absence of any extension primers).

Besides just the templates interacting with one another, the extension primers would at times interact with templates other than those intended. Figure 17 shows an attempted Duplex with TSC810 and M9 using TSC810.2E and M9E(3) extension primers. When analyzing the multiplex it was observed that within the TSC810 variable region (dispensation number 2) of the pyrogram an A signal was always observed even though the singleplex for TSC810 did not show this background signal. Through examination of the DNA strands present within the reaction, a possible alternate binding for M9E(3) was observed (Figure 17b). This alternate binding site not only accounted for the observed A signal at dispensation 2, but also the increased T peak at dispensation number 6 (figure 17a). The T peak at dispensation number 6 should be equivalent to

that of the T peak at dispensation number 11, but is routinely about 1.2 times higher at dispensation number 6 than dispensation number 11. These background signals were eliminated by altering the extension primers or by changing which loci were examined in the same reaction, but due to the overall complications in having so many biotinylated templates present in the pyrosequencing reaction, alternate methods of multiplexing were examined.

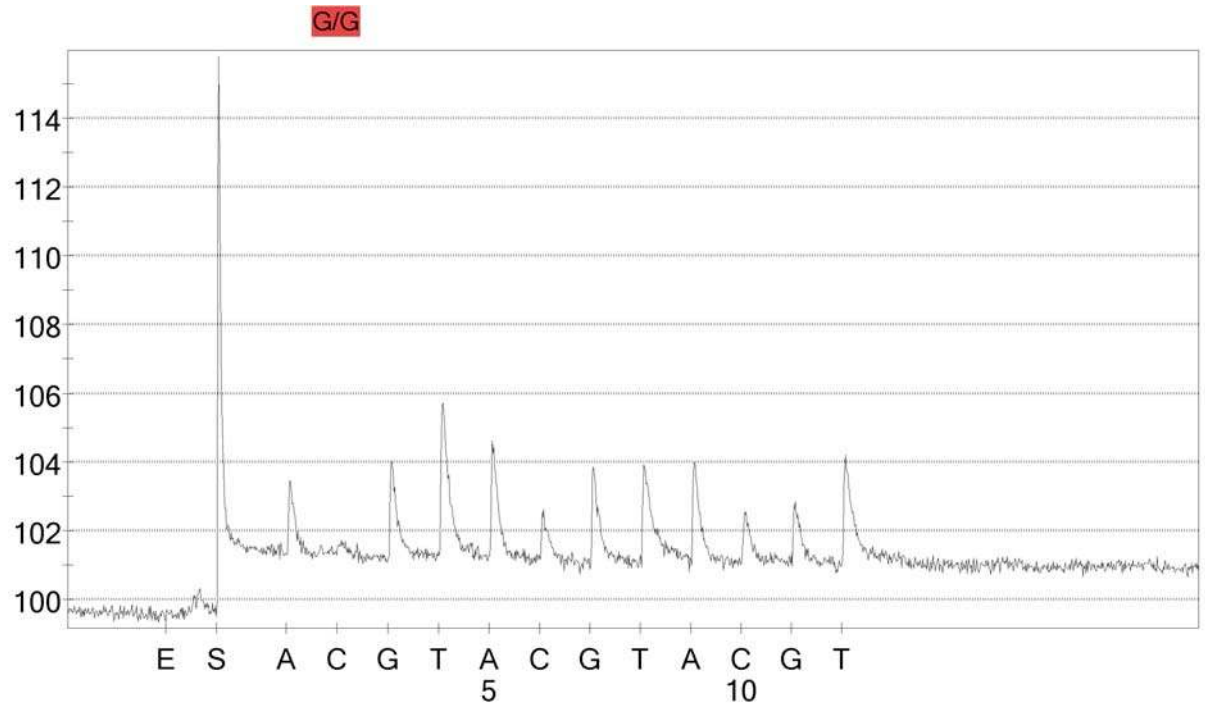


Figure 16. Template-Template Binding. Male sample (C7), pyrosequenced using only the amplified template of the M17/M3/RS640 triplex. Template only sample was analyzed using a sequential dispensation of all nucleotides.

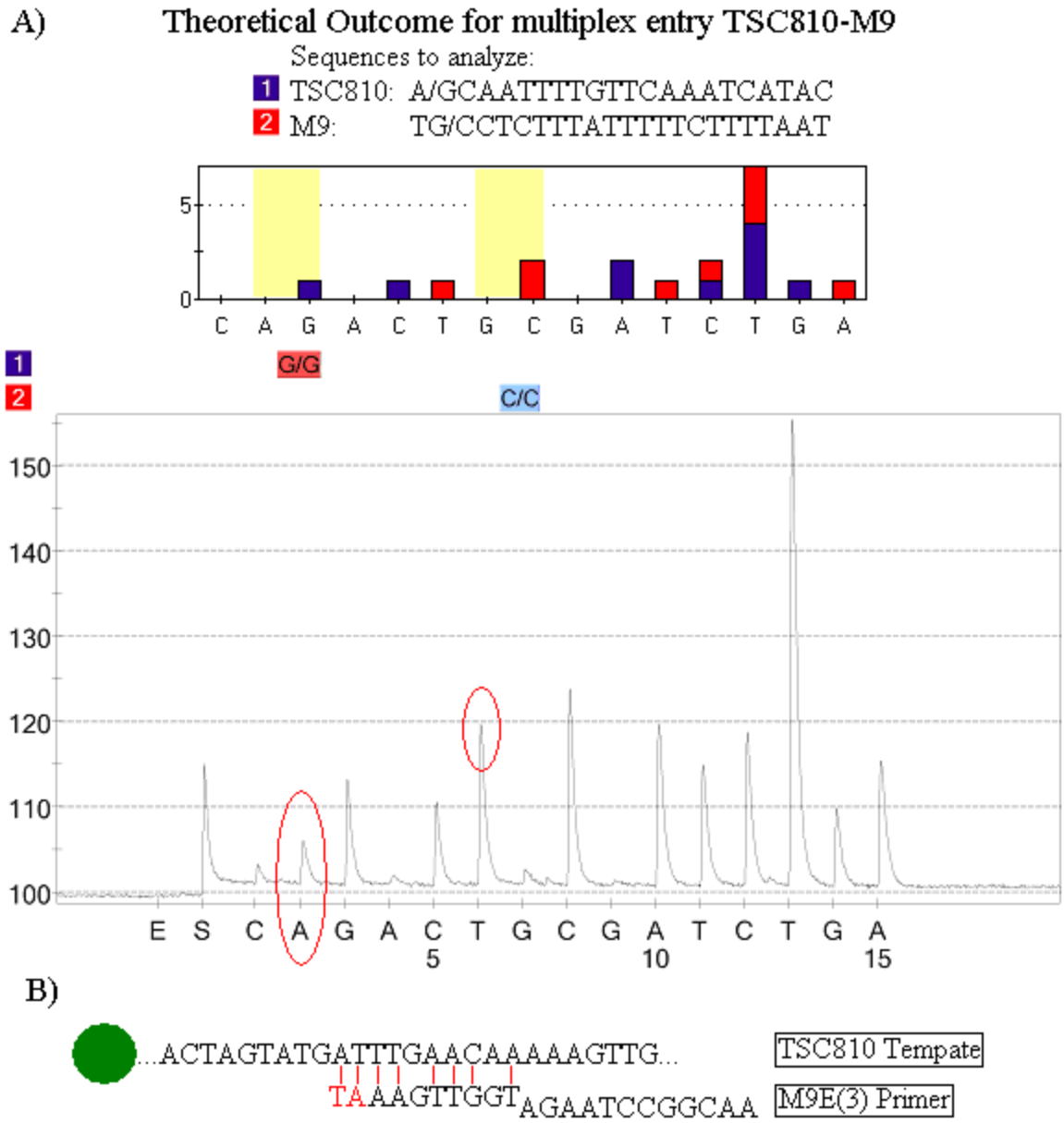


Figure 17. Background in Multiplex. A) Male individual (ME) at TSC810/M9 Duplex using TSC810.2E and M9E(3) as the extension primers. Red circles indicate background signals that are present in repeated samples. B) Proposed alternate binding site that leads to the incorporation of an A and T nucleotide in TSC810/M9 multiplex. Black letters indicate those belonging to DNA strands. Red letters indicate proposed incorporated nucleotides during pyrosequencing.

The nested PCR approach allowed for amplification of numerous loci while eliminating the problems observed with the high number of templates, by utilizing only three biotinylated primer sets in the second reaction. Approximately 90 different Duplex and Triplexes PCR and pyrosequencing reactions were analyzed for possible use in the nested PCR approach. The multiplexes with promising results produced no significant background patterns caused by primer-primer, primer-template or template-template interactions and were designed to allow for non ambiguous allele calling (Figure 4). Although designed for unambiguous allele calling, preferential amplification caused failed pyrosequencing reactions in approximately 98% of the pyrosequencing different multiplexes attempted, but was eliminated with Pyrosequencing Multiplexes (PMP) 1, 2, and 3.

Preferential amplification, when present, could be observed in both product gels (Figure 18) and in the pyrosequencing programs (Figure 19). Figure 18 shows preferential amplification at RS640/RS653 with RS640 (288 bp product) amplifying more than RS653 (116 bp product), and M207/RS653 with M207 (347 bp product) amplifying more than RS653 (116 bp product) when amplified as duplexes compared to when amplified as two different singleplexes. Figure 19 shows the theoretical result for the duplex (M2/M213) amplification and pyrosequencing with a single male individual (C1 - 13 plex product). The pyrogram shows sequence analysis of the M2 allele, but the M213 cannot be analyzed due to an absence of signal in the M213 nucleotide dispensations (C or T nucleotide at dispensation 2 or 3, T nucleotide at dispensation 11, A nucleotide at dispensation 12, or a G nucleotide at dispensation 14). If preferential amplification

could not be eliminated by altering primer concentrations then that combination of loci was eliminated from multiplex analysis.

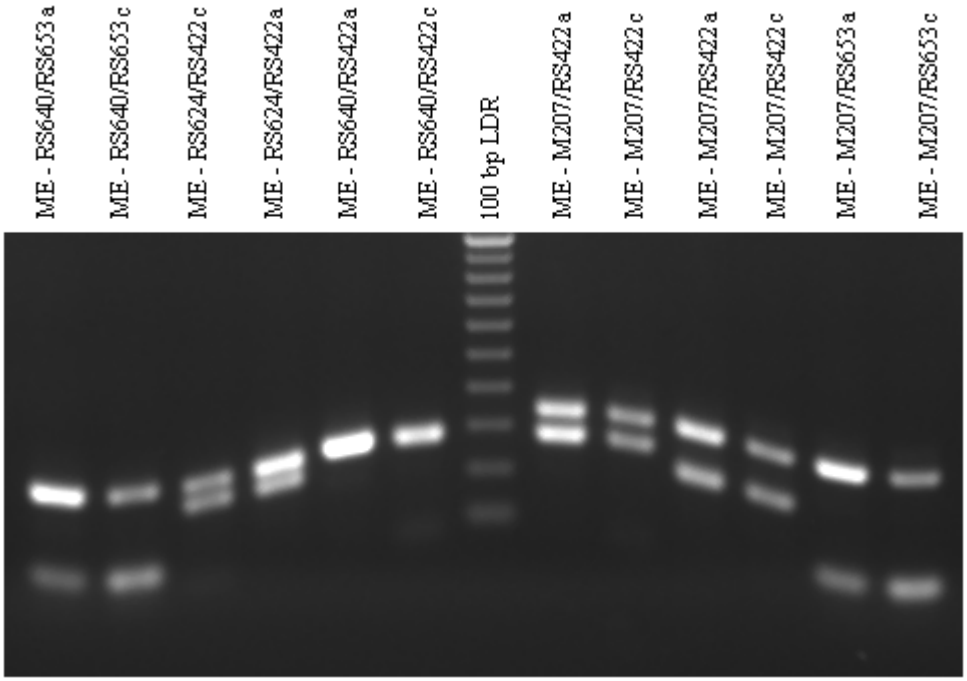


Figure 18. Preferential Amplification Gel. 1% Agarose gel containing 0.01% EtBr run at 200V for 12 min. Electrophoresis: run post amplification of male (9948 13 plex PCR product) at various duplexes. “a” in lane name represents the duplex amplification. “c” in lane name represents control where each loci was amplified individually then the PCR products mixed for purposes of running on the gel.

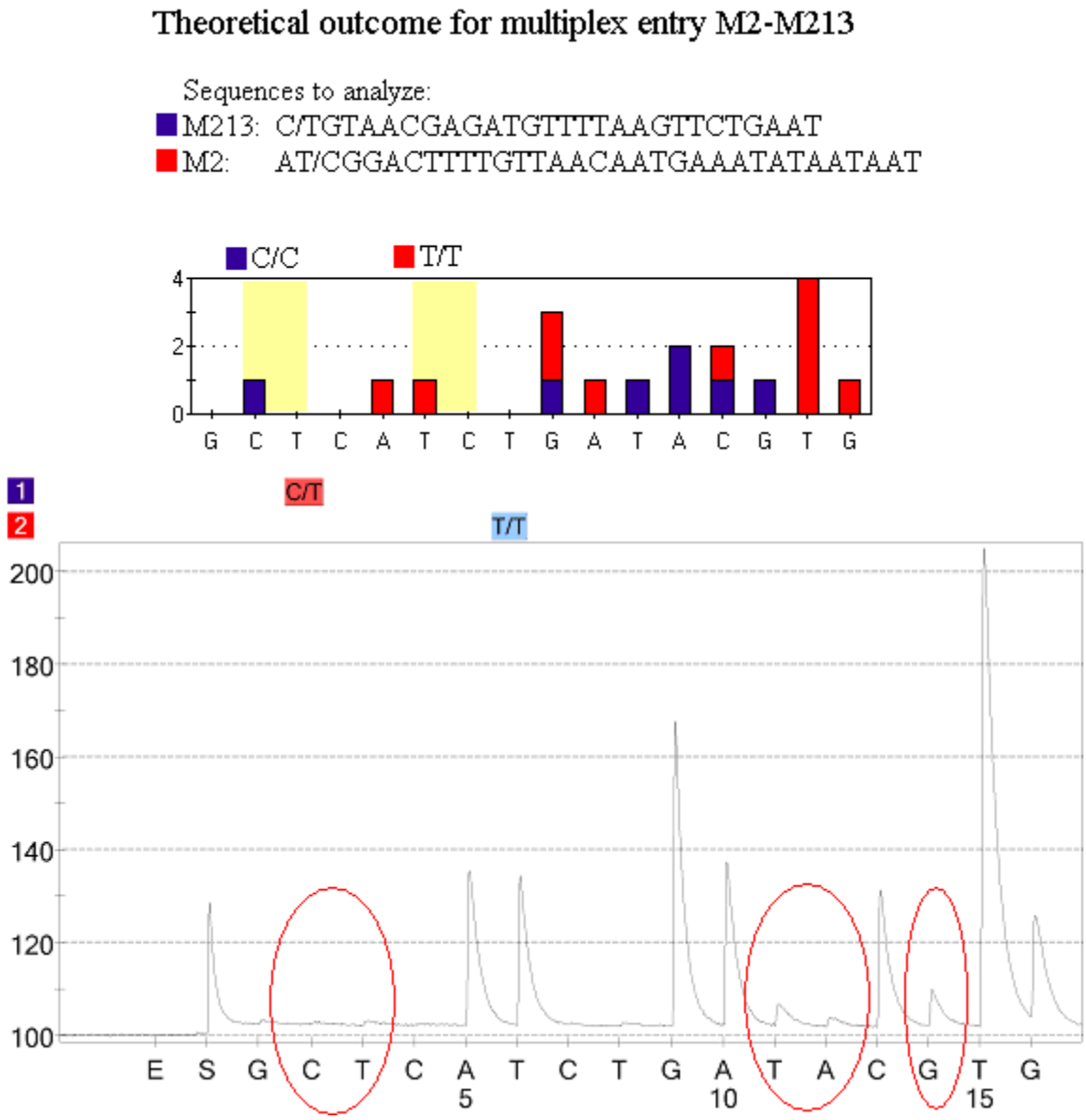


Figure 19. Preferential Amplification Pyrogram. Male individual (C1-13 plex product) amplified at M2/M213 and analyzed with M2.2E(2) and M213E. Pyrogram is missing M213 sequence (red circles) when compared with theoretical outcome (blue).

Through altering PCR primer concentrations and loci selection, three multiplexes were designed. This allowed for routine unambiguous sequencing of selected polymorphic loci: PMP1 (M207/RS843/PN2), PMP2 (TSC810/PN1/RS640), and PMP3 (M9/M170). Each one of these multiplexes allows for unambiguous allele calling, and amplification in a single reaction. PMP1 requires the multiplex PCR reaction described in chapter 2 with 8.0 μM M207 forward and reverse primers, 0.3 μM RS843 forward and reverse primers, and 0.325 μM PN2 forward and reverse primers. Figure 20 shows the dispensation order and all possible SNP patterns for single source male individuals when M207E(2), PN2E, and RS843E are used as extension primers. PMP2 uses the multiplex PCR reaction with 1.6 μM PN1 forward and reverse primers, 2.8 μM TSC810 forward and reverse primers, and 2.0 μM RS640 forward and reverse primers. Figure 21 shows the dispensation order and all possible SNP patterns for single source male individuals when TSC810E(2), PN1E, and RS640E are used as extension primers. PMPs uses the multiplex PCR reaction with 2.0 μM M9 forward and reverse primers, and 2.0 μM M170 forward and reverse primers. Figure 22 shows the dispensation order and all possible SNP patterns for single source male individuals when M170E and M9E(4) are used as extension primers.

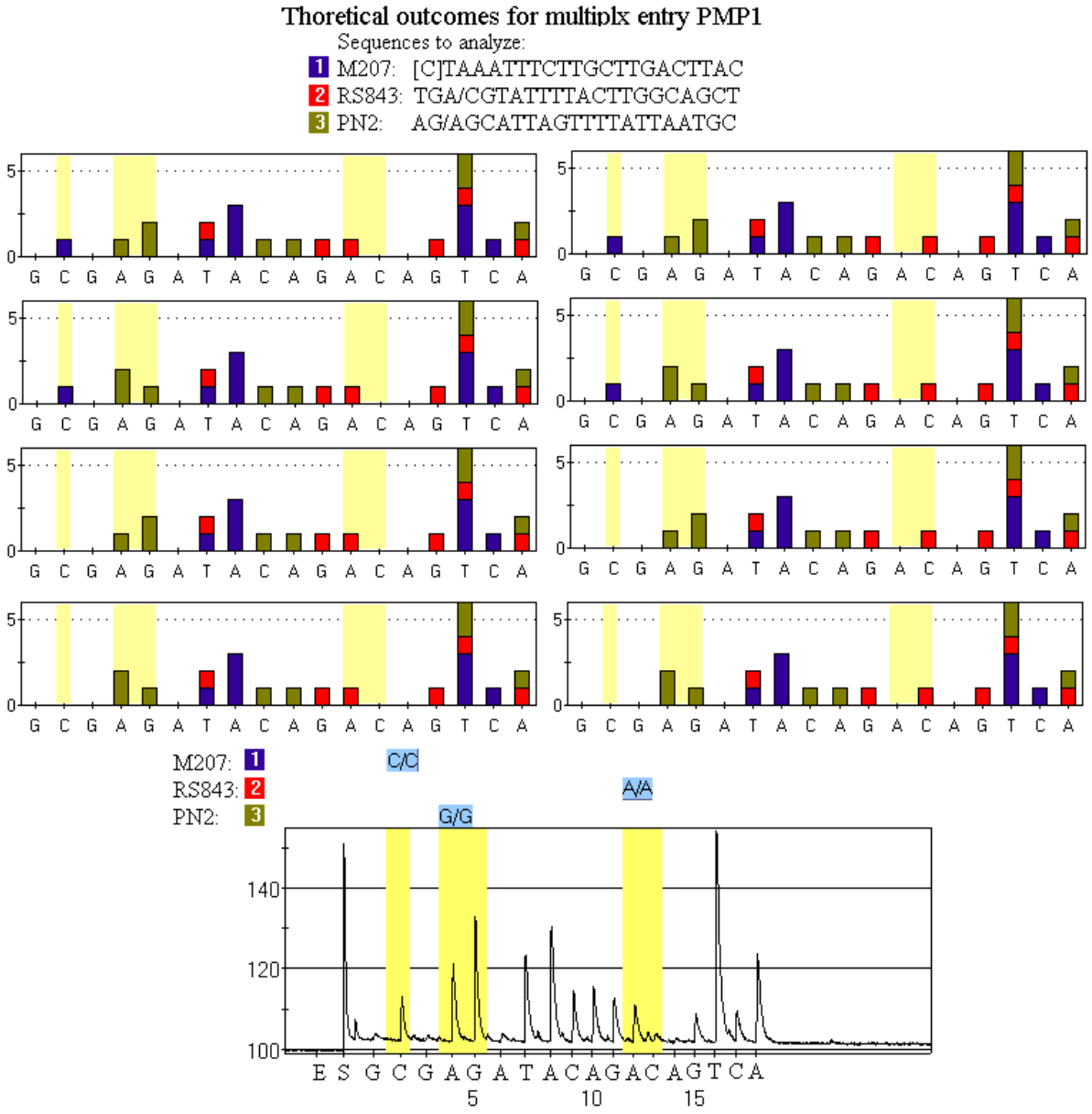


Figure 20. PMP1 Outcomes. Sequences to analyze are those when M207E(2), RS843E, and PN2E are used for analysis. Eight possible outcomes are possible. The PSQ96MA Software uses blue to reference the M207 sequence, red to indicate the RS843 sequence, and green to reference the PN2 sequence. Pyrogram is from single male source (9948) showing a C allele at M207, G allele at PN2, and an A allele at RS843.

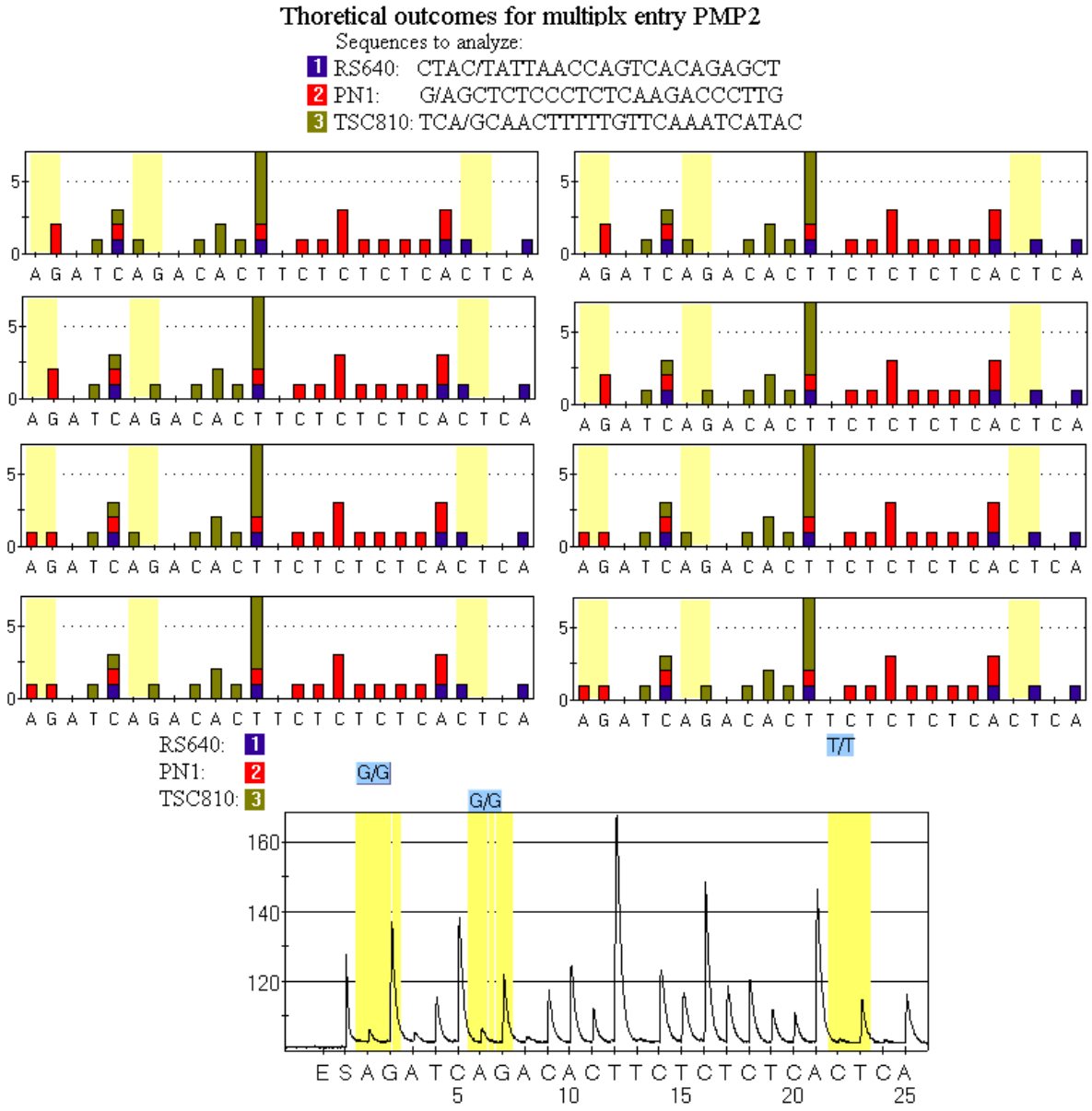


Figure 21. PMP2 Outcomes. Sequences to analyze are those when TSC810E(2), PN1E, and RS640E are used for analysis. Eight possible outcomes use blue to reference RS640 sequence, red to indicate PN1 sequence, and green to reference TSC810 sequence. Pyrogram is from single male source (ME) showing a G allele at PN1, G allele at TSC810, and a T allele at RS640.

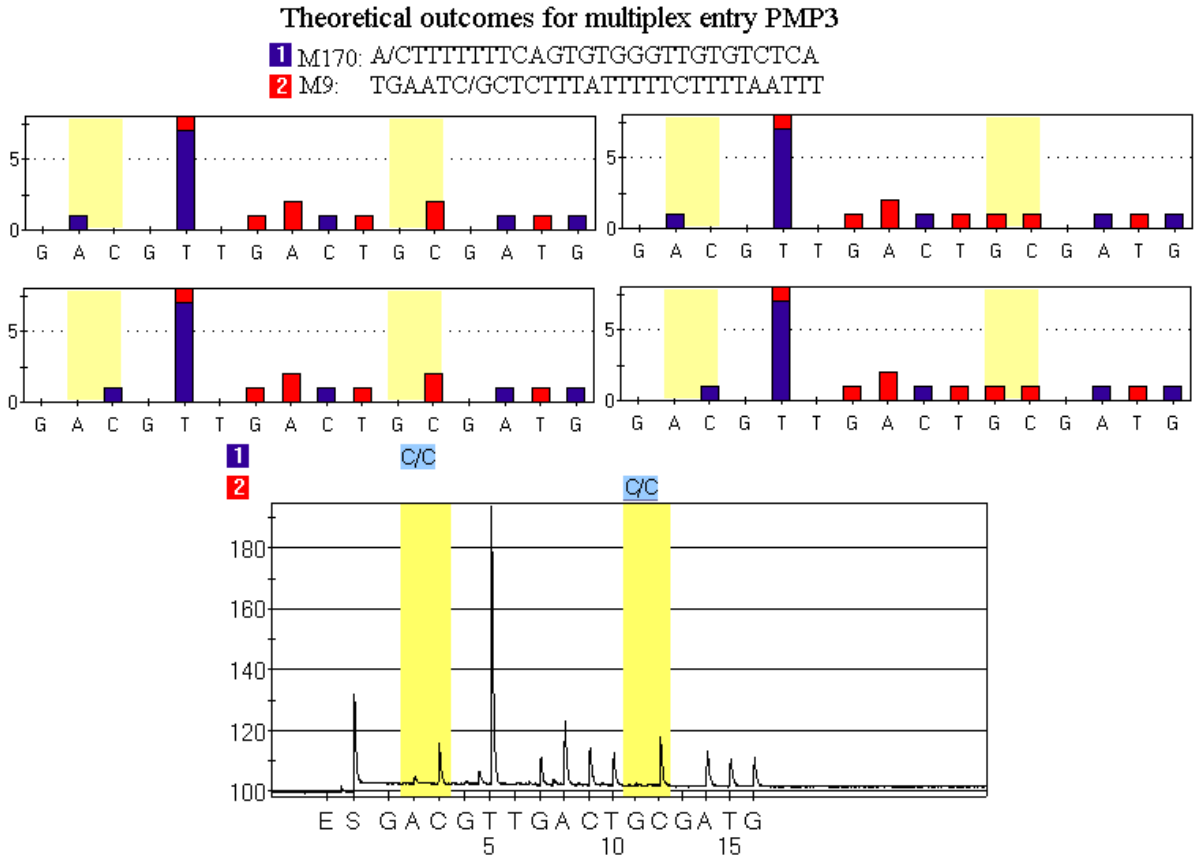


Figure 22. PMP3 Outcomes. Sequences to analyze are those when M170E and M9E(4) are used for analysis. Four possible outcomes use blue to reference M170 sequence and red to indicate M9 sequence. Pyrogram is from single male source (C7) showing a C allele at M170 and a C allele at M9.

3.5 Forensic Value Analysis

3.5.1 Loci Sequencing

In all, 50 Y chromosome loci were suitable for male specific amplification and were able to be analyzed using pyrosequencing. All 50 loci were sequenced in 60 different individuals to

determine rate of polymorphism. Individuals tested belonged to three different ethnicities, Caucasians (n = 20), African Americans (n = 18), Hispanics (n = 21), and one control (9948). Caucasian samples ME and C1 – C13, African American samples B1 and B2 and, the Hispanic sample H1 were collected with buccal swabs. Caucasian samples 174-179 and African American samples 181-196 were donated as dried blood stains. Hispanic samples 570-589 were contributed as dried down DNA extracts. These samples were amplified at all 50 loci, individually and examined using singleplex pyrosequencing so as to determine each individual's haplotype without ambiguity and determine the exact number of alleles observed for each locus with the 60 examined individuals. Table 6 lists the 50 loci examined and the alleles present for each at all 60 individuals.

3.5.2 Loci Allele Distribution

The 50 loci were analyzed across all 60 individuals to determine the rate of polymorphism observed. The loci fell into four distinct classes: highly polymorphic (major allele observed in 50% to 95% of the individuals) slightly polymorphic (major allele observed in 95% to 99% of the individuals), and non polymorphic (major allele observed in all individuals). Twenty highly polymorphic loci were observed: M207, RS407, M173, RS843, M45, M74, RS035, TSC810, P25, M9, M89, M213, RS640, RS653, RS422, PN2, P29, M2, PN1, and M170. Twelve slightly polymorphic loci were observed: M3, M17, M60, M69, M124, M168, M72, M181, M182, M214, RS155, and SRY10831. Eighteen non polymorphic loci were observed: M6, M11, M42, M63, M75, M112, M122, M123, M132, M166, M174, M203, M215, M216, M217, PN3, RS664, and 92R7.

The polymorphic loci belong in ten distinct groups. Loci within the same group have major and minor alleles that have an appearance of being genetically linked in this data set. For instance, any individual who has the deletion at M207 also has a G allele at RS407. Conversely, any individual that has the C allele at M207 also has the C allele at RS407. Table 7 lists the 10 different groups, in which the loci are grouped together, their observed major allele combination, and their major allele frequency for all 59 individuals (9948, male control, was excluded in allele frequency calculations).

The 12 slightly polymorphic loci (M3, M17, M60, M69, M124, M168, M72, M181, M182, M214, RS155, and SRY10831) were all similar in that the minor allele was found in only one of the 59 individuals and controls examined. The individual minor alleles were observed in 7 different individuals: C3 (M124), C5 (M17 and M168), 176 (M172), 185 (M182, and M181), 582 (M3), 587 (M69), and 589 (M214, RS155). The 18 non-polymorphic loci exhibited the same major allele in all individuals and controls examined.

A closer look at each highly polymorphic locus group and their allele distribution shows significant racial differences. For 9 of the 10 highly polymorphic locus groups (groups 1 – 9) the major allele in the Caucasian population is the minor allele in the African American population and vice versa. The Hispanic racial distribution does not show a preference for associating with either the African American or Caucasian racial groups.

Table 6

Loci Allele Designations for All Individuals

Sample	Race	M207	RS407	M173	RS843	M45	M74	TSC810	RS035	M9	RS640	M89	M213	RS653	RS422	PN2	P29	PN1	M2	M170	M17	M182	M181
181	African American	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
182	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
183	African American	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
184	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
185	African American	del	G	T	C	C	C	G	G	C	C	G	T	C	T	G	A	G	T	A	G	T	C
186	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
187	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
188	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
189	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
190	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
191	African American	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	A	G	C	T
192	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
193	African American	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
194	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
195	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
196	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
B1	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
B2	African American	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
C1	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C2	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C3	Caucasian	C	C	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C4	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C5	Caucasian	C	C	G	A	T	T	G	A	G	T	A	C	C	T	G	A	G	T	A	T	C	T
C6	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C7	Caucasian	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
C8	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C9	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C10	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C11	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
C12	Caucasian	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	A	G	C	T
C13	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T

Sample	SRY 10831	M168	M69	M3	M124	M172	M214	M60	RS155	PN3	M75	M132	M215	M112	M174	M42	M122	M123	M166	M203	M6	M63	RS664	M216	M217	92R7	P25
181	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
182	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
183	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
184	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
185	C	G	T	C	G	T	T	A	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
186	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
187	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
188	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
189	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
190	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
191	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
192	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
193	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
194	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
195	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
196	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
B1	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
B2	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
C1	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C2	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C3	C	A	T	C	A	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
C4	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C5	T	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
C6	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C7	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
C8	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C9	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C10	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C11	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
C12	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
C13	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG

Sample	Race	M207	RS407	M173	RS843	M45	M74	TSC810	RS035	M9	RS640	M89	M213	RS653	RS422	PN2	P29	PN1	M2	M170	M17	M182	M181
174	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
175	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
176	Caucasian	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	A	G	C	T
177	Caucasian	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
178	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
179	Caucasian	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
ME	Caucasian	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
570	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
571	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
572	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
573	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
574	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
575	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
576	SW Hispanic	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
577	SW Hispanic	del	G	T	C	C	C	G	G	G	T	A	C	C	T	G	A	G	T	A	G	C	T
578	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
579	SW Hispanic	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
580	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
581	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
582	SW Hispanic	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
583	SW Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
584	SW Hispanic	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	A	G	C	T
585	SW Hispanic	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
586	SW Hispanic	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	C	G	C	T
587	SW Hispanic	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	G	T	A	G	C	T
588	SW Hispanic	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
589	SW Hispanic	del	G	T	C	C	C	G	G	G	T	A	C	C	T	G	A	G	T	A	G	C	T
H1	SE Hispanic	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T
9948	Control (Male)	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	A	G	C	T

Sample	SRY 10831	M168	M69	M3	M124	M172	M214	M60	RS155	PN3	M75	M132	M215	M112	M174	M42	M122	M123	M166	M203	M6	M63	RS664	M216	M217	92R7	P25
174	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
175	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
176	C	A	T	C	G	G	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
177	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
178	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
179	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
ME	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
570	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
571	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
572	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
573	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
574	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
575	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
576	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
577	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
578	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
579	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
580	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
581	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
582	C	A	T	T	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
583	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
584	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
585	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
586	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
587	C	A	C	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
588	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
589	C	A	T	C	G	T	C	del	C	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G
H1	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	T/GG
9948	C	A	T	C	G	T	T	del	T	C	G	G	T	C	T	T	A	C	C	G	T	G	T	G	T	G	G

"Sample" are local identifiers supplied by the NCFS laboratory. "Race" refers to the ethnic origin of the individual who supplied the sample as either "Caucasian", "African American", or "Hispanic". Alleles A,G,T,C are those observed using pyrosequencing and would be complimentary to the template sequence (biotin labeled PCR primer strand). "del" indicates a deletion of the expected nucleotide at that examined position.

Table 7

Polymorphic Loci Groups

Group	Loci	Observed Major Allele Combination	Major Allele (n=59)
1	M207, RS407	del, G	51%
2	M173, RS843	T, G	53%
3	M45, M74, RS035	T, T, A	54%
4	TSC810	G	54%
5	M9	G	58%
6	P25	G	59%
7	M89, M213, RS640	A, C, T	73%
8	RS653, RS422, PN2, P29	C, T, G, A	75%
9	M2, PN1	T, G	76%
10	M170	A	90%

"Observed Major Allele Combination" is the alleles in order of listed loci that appear most frequently in the examined population and segregate together within a group. "Major Allele" is the percentage of examined individuals who posses the major allele combination.

This finding correlates with the fact that Y chromosome SNPs are currently being employed to differentiate between various ethnic geographic origins similar to the Caucasian, African American, and Hispanic populations used. It also provides evidence that forensic Y chromosome SNP analysis may be of benefit in racial determinations, when the origin of a sample is unknown.

Table 8

Racial Distribution

Group	Loci	Observed Major Allele Combination	Caucasian (n=20)	African American (n=18)	Hispanics (n=21)
1	M207, RS407	del, G	25%	83%	48%
2	M173, RS843	T, G	30%	83%	48%
3	M45, M74, RS035	T, T, A	75%	17%	67%
4	TSC810	G	35%	83%	48%
5	M9	G	75%	17%	76%
6	P25	G	35%	89%	57%
7	M89, M213, RS640	A, C, T	100%	22%	90%
8	RS653, RS422, PN2, P29	C, T, G, A	100%	28%	90%
9	M2, PN1	T, G	100%	28%	95%
10	M170	A	85%	100%	86%

"Observed Major Allele Combination" is the alleles in order of listed loci that appear most frequently in the examined population and are linked within a group. n is the number of individuals in each racial class.

3.5.3 Haplogroups

The 60 different individuals and controls were examined to determine the number of distinct haplogroups that would be created using the 20 highly polymorphic and the 12 slightly polymorphic loci. Through comparative analysis, 15 different haplogroups were produced using the 60 different individuals and controls examined. Table 9 lists the 60 individuals and places them into haplogroups based on their SNP results. Each color is its own haplogroup, except for the individuals designated by white backgrounds, these 8 individuals have unique SNP results and therefore belong to their own distinct haplogroup. The largest haplogroup (shown in purple

in Table 9) contains 22 of the 60 individuals, or approximately 37% of the individuals examined. The 22 individuals came from all three of the racial groups: 11 Caucasians, 2 African Americans, and 9 Hispanics. The second largest

Table 9

Haplogroups

Sample	M207	RS407	M173	RS843	M45	M74	TSC 810	RS035	M9	RS640	M89	M213	RS653	RS422	PN2	P29	PN1	M2	P25	M170	M17	M182	M181	SRY10831	M168	M69	M3	M124	M172	M214	M60	RS155
C7	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	C	G	C	T	C	A	T	C	G	T	T	del	T
576	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	C	G	C	T	C	A	T	C	G	T	T	del	T
585	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	C	G	C	T	C	A	T	C	G	T	T	del	T
586	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	C	G	C	T	C	A	T	C	G	T	T	del	T
191	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
C12	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
177	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	T	C	G	C	T	C	A	T	C	G	T	T	del	T
ME	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	T	C	G	C	T	C	A	T	C	G	T	T	del	T
9948	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
183	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
175	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
179	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
573	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
578	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
C5	C	C	G	A	T	T	G	A	G	T	A	C	C	T	G	A	G	T	G	A	T	C	T	T	A	T	C	G	T	T	del	T
C3	C	C	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	A	T	T	del	T
587	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	G	T	G	A	G	C	T	C	A	C	C	G	T	T	del	T
185	del	G	T	C	C	C	G	G	C	C	G	T	C	T	G	A	G	T	G	A	G	T	C	C	G	T	C	G	T	T	A	T
176	del	G	T	C	C	C	G	G	C	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	G	T	del	T
589	del	G	T	C	C	C	G	G	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	C	del	C
577	del	G	T	C	C	C	G	G	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T	del	T
582	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	T	G	T	T	del	T

580	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
581	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
583	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C1	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C10	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C11	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C13	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C2	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C4	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C6	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C8	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
C9	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T
H1	C	C	G	A	T	T	A	A	G	T	A	C	C	T	G	A	G	T	T	A	G	C	T	C	A	T	C	G	T	T

579	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T
588	del	G	T	C	T	T	G	A	G	T	A	C	C	T	G	A	G	T	G	A	G	C	T	C	A	T	C	G	T	T

182	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
184	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
186	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
187	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
188	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
189	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
190	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
192	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
194	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
195	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
196	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
B1	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
B2	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T
584	del	G	T	C	C	C	G	G	C	C	G	T	G	A	A	C	A	C	G	A	G	C	T	C	A	T	C	G	T	T

Each color represents one of the 15 unique haplogroups. Individuals within the same color belong to the same haplogroup, except for white. Individuals in white belong separately to their own distinct haplogroup.

haplogroup (shown in orange in Table 9) contained 14 individuals (approximately 23% of the individuals examined) of which 13 were African American and 1 was Hispanic; no Caucasians were found within this haplogroup. There were 5 haplogroups that contained between 2 and 6 individuals.

Through examination of the loci results and haplogroups, it was determined that not all of the 20 highly polymorphic and 12 slightly polymorphic loci are required to differentiate the 60 individuals into the 15 different haplogroups. Only 13 of the 32 loci are required: one loci from each of the 10 highly polymorphic loci groups, M3, M172, and M214. Table 10 shows that using only these 13 loci, the 60 individuals are broken down into the same 15 haplogroups.

3.5.4 Various DNA Sources

To determine if the 13plex pyrosequencing analysis system could be successful with most common DNA sources found in forensic examinations, epithelial cells, blood, semen, and saliva DNA from the same male individual (ME) was analyzed using the PMP1, PMP2, PMP3 multiplexes and M3, M74, M172, M214, and P25 simplexes. All four DNA sources produced unambiguous pyrosequencing results for all 13 loci. Figure 23 shows the PMP1,PMP2, and PMP3 results for the semen sample and Figure 24 shows the M3, M74, M172, M214, and P25 simplex results from the saliva samples.

Table 10

Haplogroups, Minimal Loci

Sample	M207	RS843	M74	TSC810	M9	RS640	PN2	PN1	P25	M170	M3	M172	M214
174	C	A	T	A	G	T	G	G	T	A	C	T	T
178	C	A	T	A	G	T	G	G	T	A	C	T	T
181	C	A	T	A	G	T	G	G	T	A	C	T	T
193	C	A	T	A	G	T	G	G	T	A	C	T	T
570	C	A	T	A	G	T	G	G	T	A	C	T	T
571	C	A	T	A	G	T	G	G	T	A	C	T	T
572	C	A	T	A	G	T	G	G	T	A	C	T	T
574	C	A	T	A	G	T	G	G	T	A	C	T	T
575	C	A	T	A	G	T	G	G	T	A	C	T	T
580	C	A	T	A	G	T	G	G	T	A	C	T	T
581	C	A	T	A	G	T	G	G	T	A	C	T	T
583	C	A	T	A	G	T	G	G	T	A	C	T	T
C1	C	A	T	A	G	T	G	G	T	A	C	T	T
C10	C	A	T	A	G	T	G	G	T	A	C	T	T
C11	C	A	T	A	G	T	G	G	T	A	C	T	T
C13	C	A	T	A	G	T	G	G	T	A	C	T	T
C2	C	A	T	A	G	T	G	G	T	A	C	T	T
C4	C	A	T	A	G	T	G	G	T	A	C	T	T
C6	C	A	T	A	G	T	G	G	T	A	C	T	T
C8	C	A	T	A	G	T	G	G	T	A	C	T	T
C9	C	A	T	A	G	T	G	G	T	A	C	T	T
H1	C	A	T	A	G	T	G	G	T	A	C	T	T
9948	C	A	T	A	G	T	G	G	G	A	C	T	T
183	C	A	T	A	G	T	G	G	G	A	C	T	T
175	C	A	T	A	G	T	G	G	G	A	C	T	T
179	C	A	T	A	G	T	G	G	G	A	C	T	T
573	C	A	T	A	G	T	G	G	G	A	C	T	T
578	C	A	T	A	G	T	G	G	G	A	C	T	T
C7	del	C	C	G	C	T	G	G	G	C	C	T	T
576	del	C	C	G	C	T	G	G	G	C	C	T	T
585	del	C	C	G	C	T	G	G	G	C	C	T	T
586	del	C	C	G	C	T	G	G	G	C	C	T	T

Sample	M207	RS843	M74	TSC810	M9	RS640	PN2	PN1	P25	M170	M3	M172	M214
191	del	C	C	G	C	T	G	G	G	A	C	T	T
C12	del	C	C	G	C	T	G	G	G	A	C	T	T
177	del	C	C	G	C	T	G	G	T	C	C	T	T
ME	del	C	C	G	C	T	G	G	T	C	C	T	T
579	del	C	T	G	G	T	G	G	G	A	C	T	T
588	del	C	T	G	G	T	G	G	G	A	C	T	T
182	del	C	C	G	C	C	A	A	G	A	C	T	T
184	del	C	C	G	C	C	A	A	G	A	C	T	T
186	del	C	C	G	C	C	A	A	G	A	C	T	T
187	del	C	C	G	C	C	A	A	G	A	C	T	T
188	del	C	C	G	C	C	A	A	G	A	C	T	T
189	del	C	C	G	C	C	A	A	G	A	C	T	T
190	del	C	C	G	C	C	A	A	G	A	C	T	T
192	del	C	C	G	C	C	A	A	G	A	C	T	T
194	del	C	C	G	C	C	A	A	G	A	C	T	T
195	del	C	C	G	C	C	A	A	G	A	C	T	T
196	del	C	C	G	C	C	A	A	G	A	C	T	T
B1	del	C	C	G	C	C	A	A	G	A	C	T	T
B2	del	C	C	G	C	C	A	A	G	A	C	T	T
584	del	C	C	G	C	C	A	A	G	A	C	T	T
C5	C	A	T	G	G	T	G	G	G	A	C	T	T
C3	C	C	T	G	G	T	G	G	G	A	C	T	T
587	del	C	C	G	C	C	A	G	G	A	C	T	T
185	del	C	C	G	C	C	G	G	G	A	C	T	T
176	del	C	C	G	C	T	G	G	G	A	C	G	T
589	del	C	C	G	G	T	G	G	G	A	C	T	C
577	del	C	C	G	G	T	G	G	G	A	C	T	T
582	del	C	T	G	G	T	G	G	G	A	T	T	T

Each color represents one of the 15 unique haplogroups. Individuals within the same color belong to the same haplogroup, except for white. Individuals in white belong separately to their own distinct haplogroup.

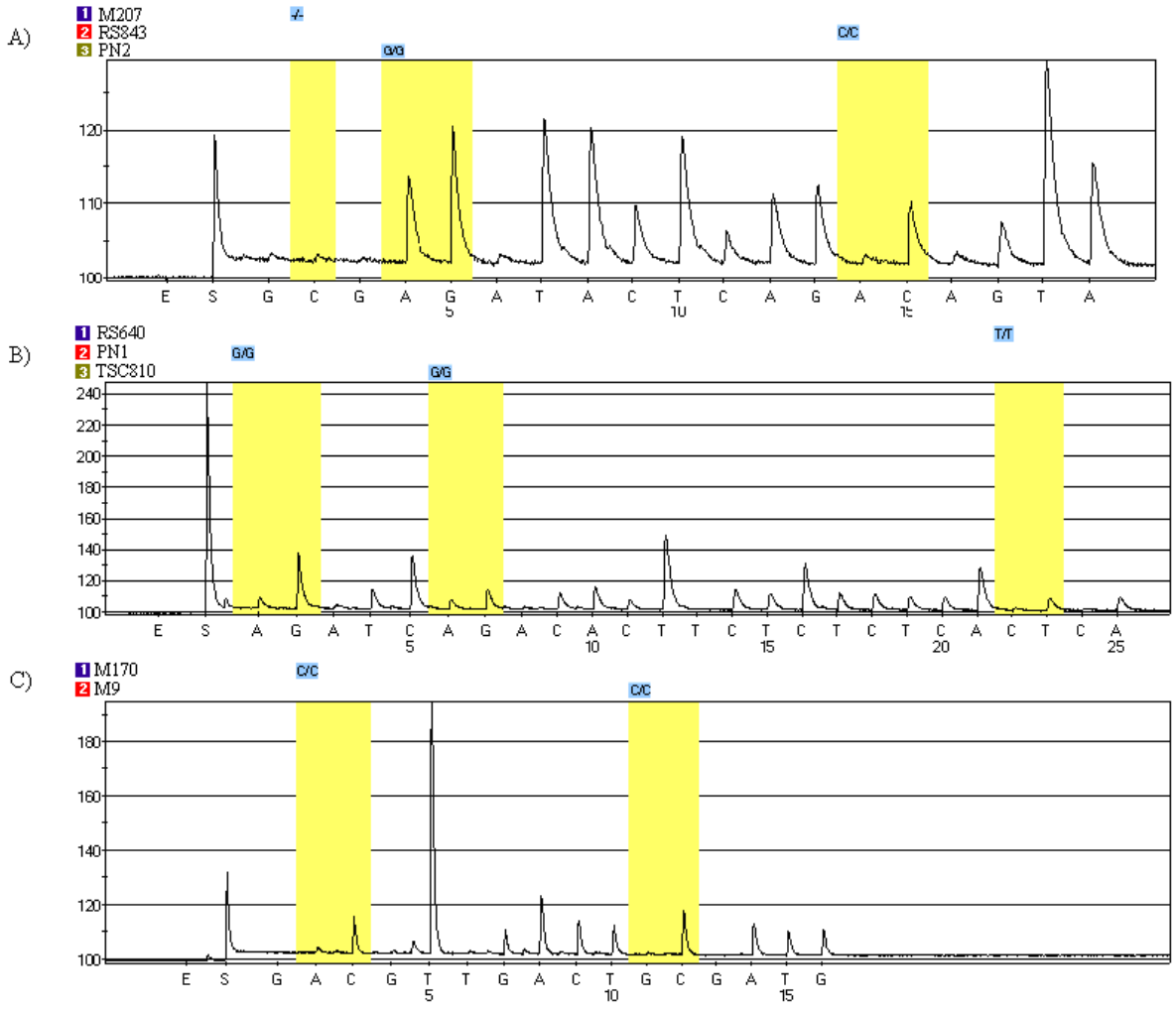


Figure 23. Semen at PMP1, PMP2, and PMP3. Male individual ME (semen) amplified with 13-plex outer primer set, and then analyzed using PMP1, PMP2, and PMP3 amplification and pyrosequencing. A) Pyrogram produced for PMP1 showing the unambiguous calling of the deletion at M207, the G allele at PN2, and the C allele at RS843. B) Pyrogram produced for PMP2 showing the unambiguous calling of the G allele at TSC810, the G allele at PN1, and the T allele at RS640. C) Pyrogram produced for PMP3 showing the unambiguous calling of the C allele at M170, and the C allele at M9.

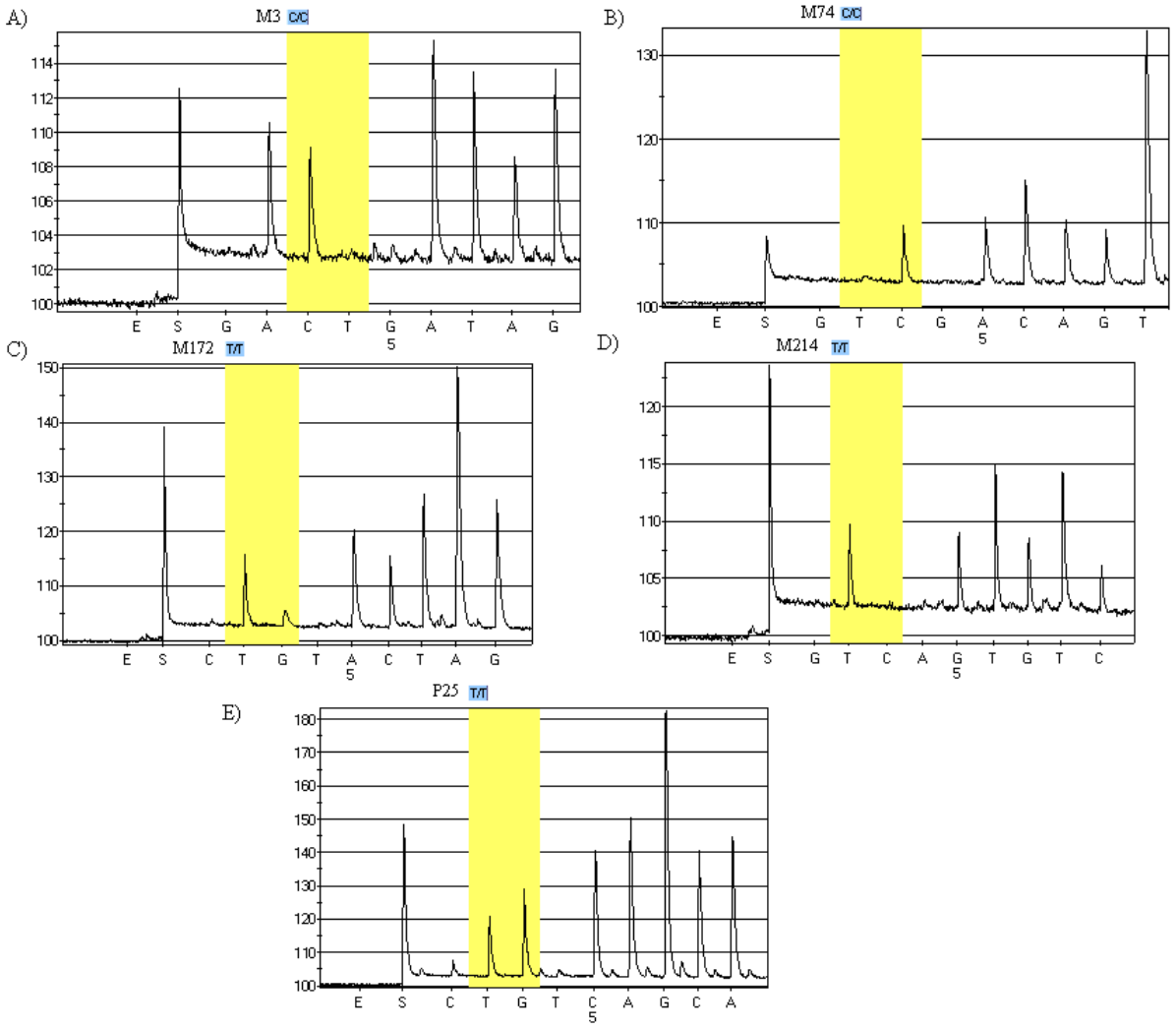


Figure 24. Saliva at M3, M74, M172, M207, and P25. Male individual ME (saliva) amplified with 13-plex outer primer set, and then analyzed using M3, M74, M172, M214, and P25 simplex amplification and pyrosequencing. A) Pyrogram produced for M3 showing unambiguous allele calling of the C allele versus the alternate T allele. B) Pyrogram produced for M74 showing unambiguous allele calling of the C allele versus the alternate T allele. C) Pyrogram produced for M170 showing unambiguous allele calling of the TTT homopolymer versus the alternate TTG sequence. D) Pyrogram produced for M214 showing unambiguous allele calling of the T allele versus the alternate C allele. E) Pyrogram produced for P25 showing unambiguous allele calling of the GGG homopolymer versus the alternate TGG sequence.

3.5.5 Minimal Starting DNA

A valuable property of forensic DNA analysis systems is their ability to be examined with small amounts of starting DNA. The starting DNA amounts tested were: 1ng, 500pg, 100pg, 50pg, 25pg, 10pg, 5pg, 1pg, and no starting DNA. Figure 25, shows a yield gel with the 1ng/μL B1 sample that was used as the starting stock to make the male DNA titrations. It was determined that a full profile (able to make allele calls for all 13 loci) was accomplished using all the titrated starting amounts of DNA, except for the 1pg and no starting DNA samples. Figure 26 shows the PMP1, PMP2, and PMP3 pyrograms for the 5 pg of starting male DNA. Figure 27 shows the M3, M74, M172, M214, and P25 pyrograms for the 5pg of starting male DNA Using the 1pg sample, PN2, TSC810, PN1, M170, and the M9 alleles were able to be determined unambiguously, figure 28 shows the PMP1, PMP2, and PMP3 pyrograms for the 1pg of starting male DNA. All other loci, (M207, RS843, RS640, M3, M74, M172, M214, and P25) appeared blank. All the samples sequenced from the no starting DNA were blank and all the alleles identified matched to the expected profile for the B1 individual utilized.

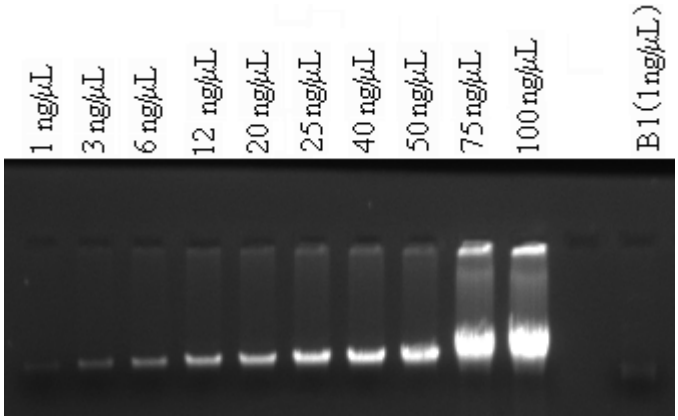


Figure 25. Dilution Yield Gel. Yield gel showing male individual B1 at 1ng/µL as compared with the DNA concentration standards. This sample was latter used for making the DNA titration of starting male DNA.

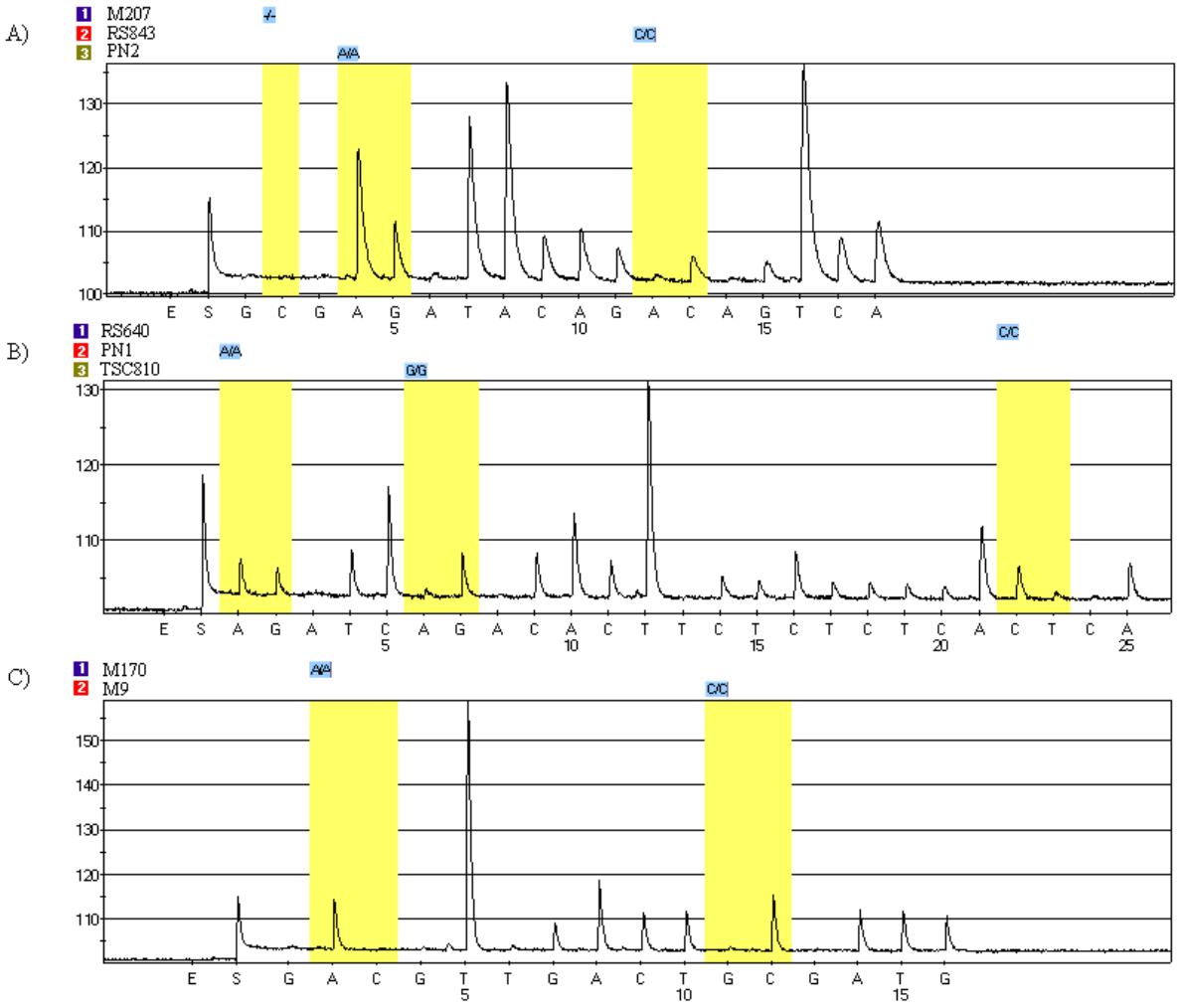


Figure 26. 5pg Male at PMP1, PMP2, and PMP3. Male individual B1 (5pg) amplified with 13-plex outer primer set, and then analyzed using PMP1, PMP2, and PMP3 amplification and pyrosequencing. A) Pyrogram produced for PMP1 showing the unambiguous calling of the deletion at M207, the A allele at PN2, and the C allele at RS843. B) Pyrogram produced for PMP2 showing the unambiguous calling of the G allele at TSC810, the A allele at PN1, and the C allele at RS640. C) Pyrogram produced for PMP3 showing the unambiguous calling of the A allele at M170, and the C allele at M9.

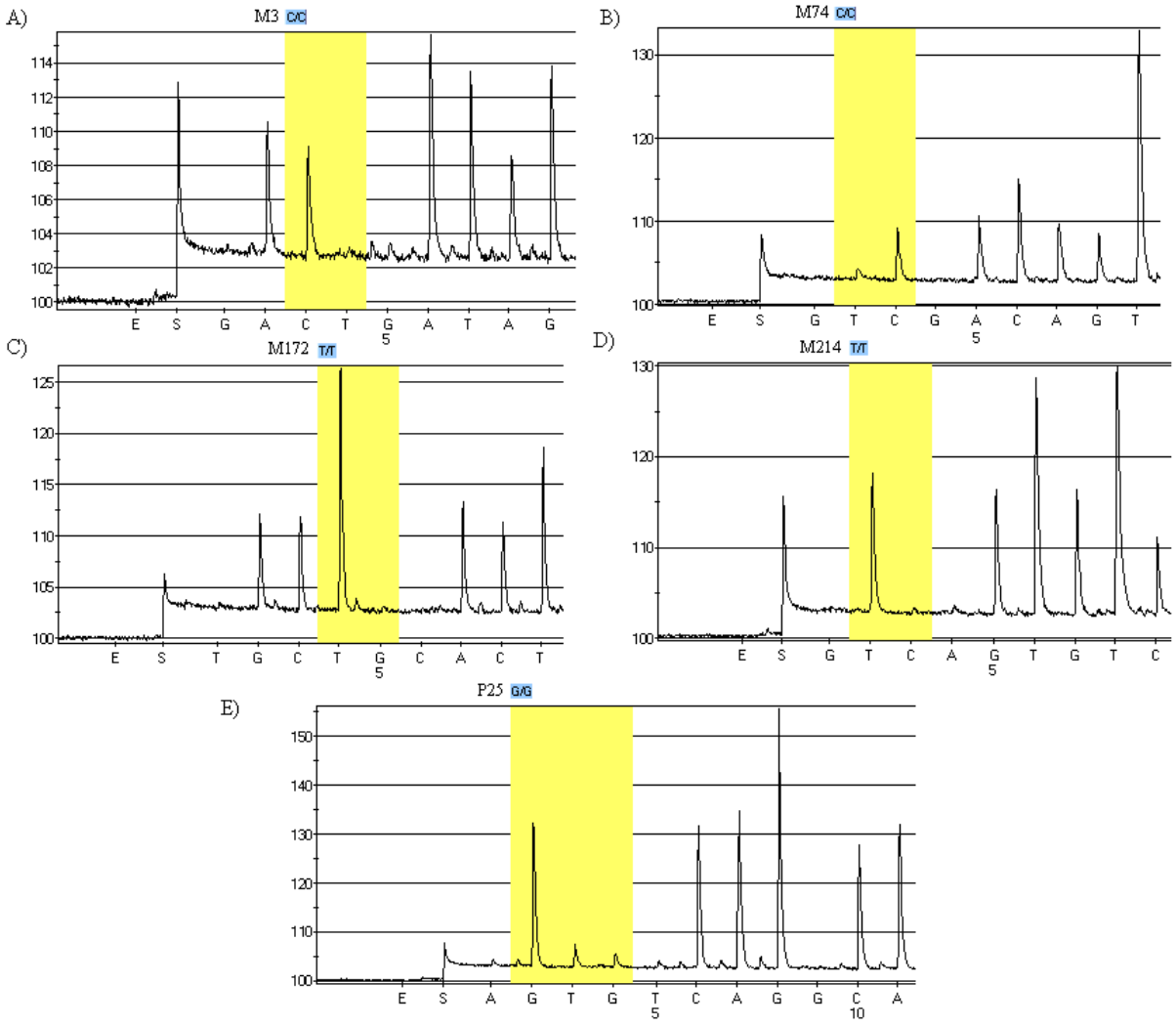


Figure 27. 5pg Male at M3, M74, M172, M207, and P25. Male individual B1 (5pg) amplified with 13-plex outer primer set, and then analyzed using M3, M74, M172, M214, and P25 simplex amplification and pyrosequencing. A) Pyrogram produced for M3 showing unambiguous allele calling of the C allele versus the alternate T allele. B) Pyrogram produced for M74 showing unambiguous allele calling of the C allele versus the alternate T allele. C) Pyrogram produced for M170 showing unambiguous allele calling of the TTT homopolymer versus the alternate TTG sequence. D) Pyrogram produced for M214 showing unambiguous allele calling of the T allele versus the alternate C allele. E) Pyrogram produced for P25 showing unambiguous allele calling of the GGG homopolymer versus the alternate TGG sequence.

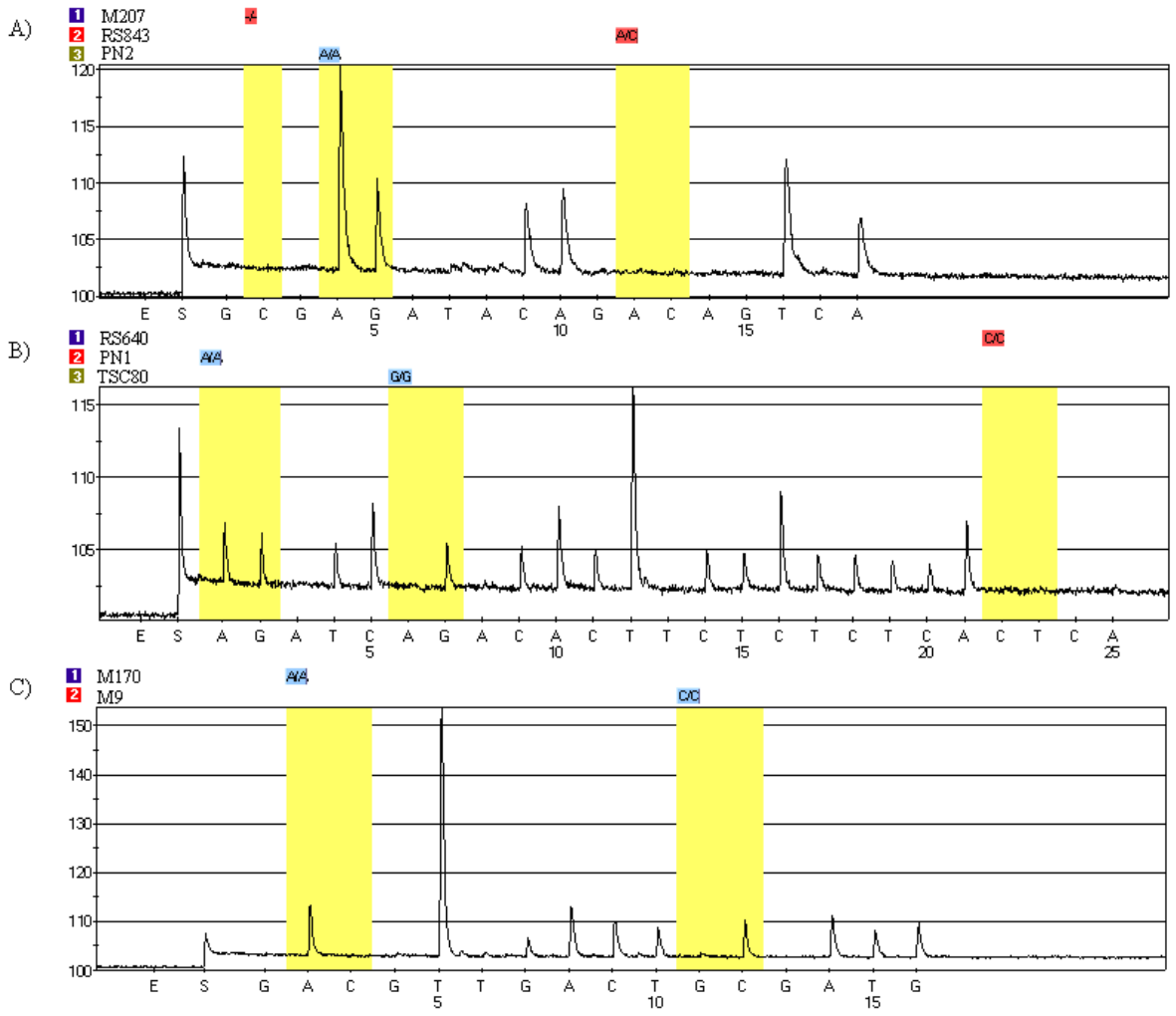


Figure 28. 1pg Male at PMP1, PMP2, and PMP3. Male individual B1 (1pg) amplified with 13-plex outer primer set, and then analyzed using PMP1, PMP2, and PMP3 amplification and pyrosequencing. A) Pyrogram produced for PMP1 showing the unambiguous calling of the A allele at PN2, and blank patterns for M207 and RS843. B) Pyrogram produced for PMP2 showing the unambiguous calling of the G allele at TSC810, the A allele at PN1, and a blank pattern for RS640. C) Pyrogram produced for PMP3 showing the unambiguous calling of the A allele at M170, and the C allele at M9.

3.5.6 Male/Female Mixture Analysis

Y chromosome analysis has many applications in forensics including human identification (down to a family of men), paternal analysis, and missing persons, but the main application is when analyzing male female mixtures. Two set of male/female mixtures were examined to determine how much female DNA could be present and still produce a full male profile with 1ng of male DNA and how much male DNA is required to get a full profile with 300ng of contaminating female DNA.

Ten different male female mixtures were analyzed: 1ng male (B1) with 8.2µg, 5µg, 1µg 500ng, and 100ng female DNA (9947a) as well as 300ng 9947a mixed with 25pg, 10pg, 300ng, 5pg, 1pg, and no male DNA (B1). It was determined that a full profile (able to make allele calls for all 13 loci) was accomplished using 1ng of B1 with all concentrations of 9947a (100ng – 8.2µg), as well as with 25pg and 10pg of B1 mixed with 300ng of 9947a. Figures 29 and 30 show the pyrograms with unambiguous allele calling produced when 10pg of male sample is mixed with 300ng of female sample. Partial profiles were accomplished with 5pg (M207, RS843, PN2, PN1, RS640, M170, M9, M3, M74, M214, P25 were called, TSC810 and M172 were blank) and 1pg (PN2, PN1, M170, M9, M3, and M214 were called, M207, RS843, TSC810, RS640, M74, M172, and P25 were blank) of B1 mixed with 300ng of 9947a.

These results indicate that the presence of female DNA has only a minor effect in the Y chromosome SNP analysis system, since with 1ng of male DNA an almost unlimited amount of female DNA can also be present and a full profile is still obtainable (the most we were able to add was 8.2 µg since we had 9947a at 300ng/µL concentration and could only add 27.2µL of

female DNA due to limitation in sample volume allowed) and 300ng of female DNA requires the addition of only 5pg more DNA than without female DNA, which is roughly equivalent to the amount of nuclear DNA found in a single epithelial or 2 sperm cells. Another interpretation of the data is that a 1 to 30,000 (10pg to 300ng) mixture of male to female DNA provides a full 13 SNP profile using the current Nested Y chromosome SNP system analyzed with pyrosequencing.

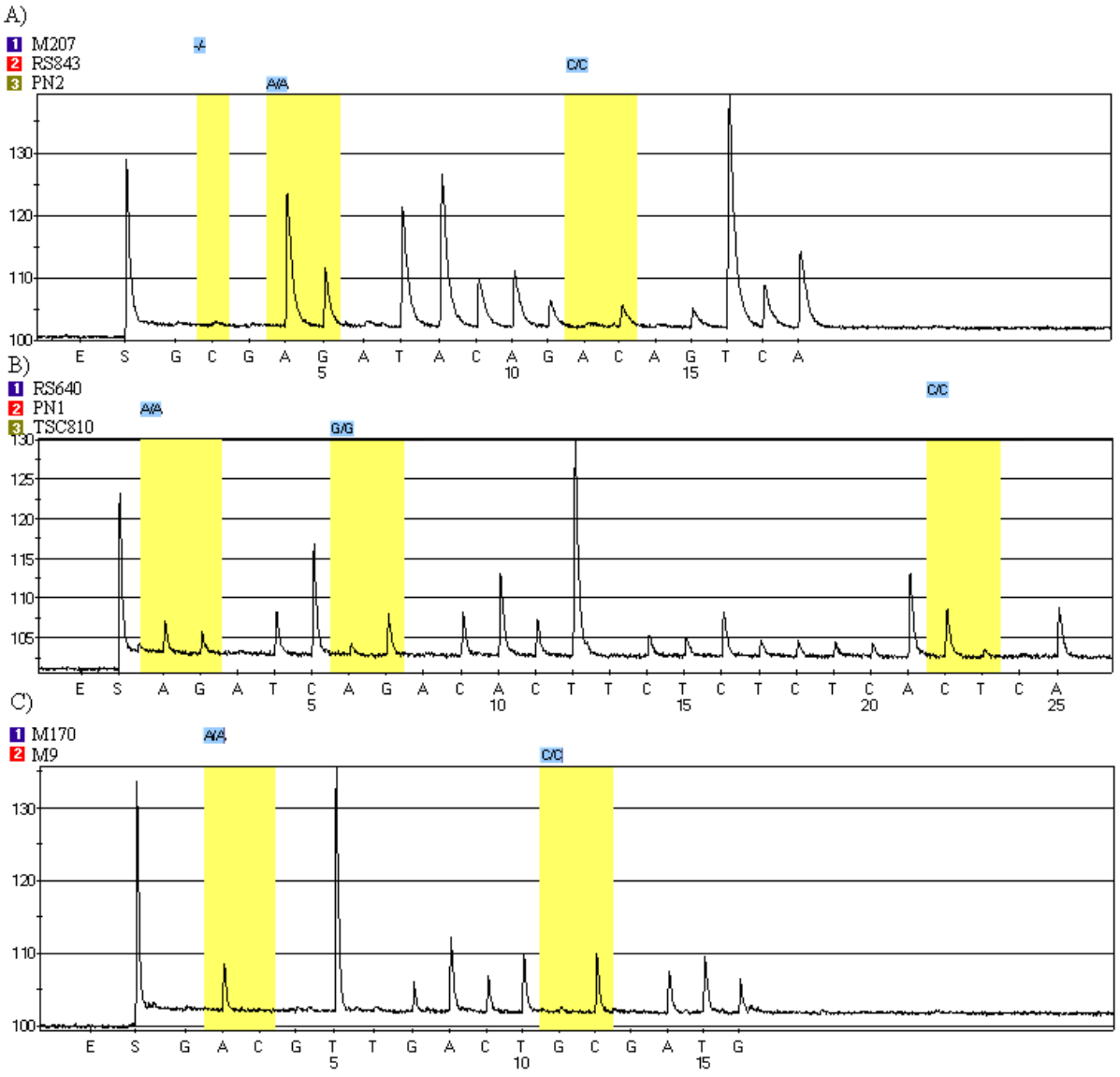


Figure 29. 10pg Male with 300ng Female at PMP1, PMP2, and PMP3. Male individual B1 (10pg) mixed with Female individual 9947a (300ng) amplified with 13-plex outer primer set, and then analyzed using PMP1, PMP2, and PMP3 amplification and pyrosequencing. A) Pyrogram produced for PMP1 showing the unambiguous calling of the deletion at M207, the A allele at PN2, and the C allele at RS843. B) Pyrogram produced for PMP2 showing the unambiguous calling of the G allele at TSC810, the A allele at PN1, and the C allele at RS640. C) Pyrogram produced for PMP3 showing the unambiguous calling of the A allele at M170, and the C allele at M9.

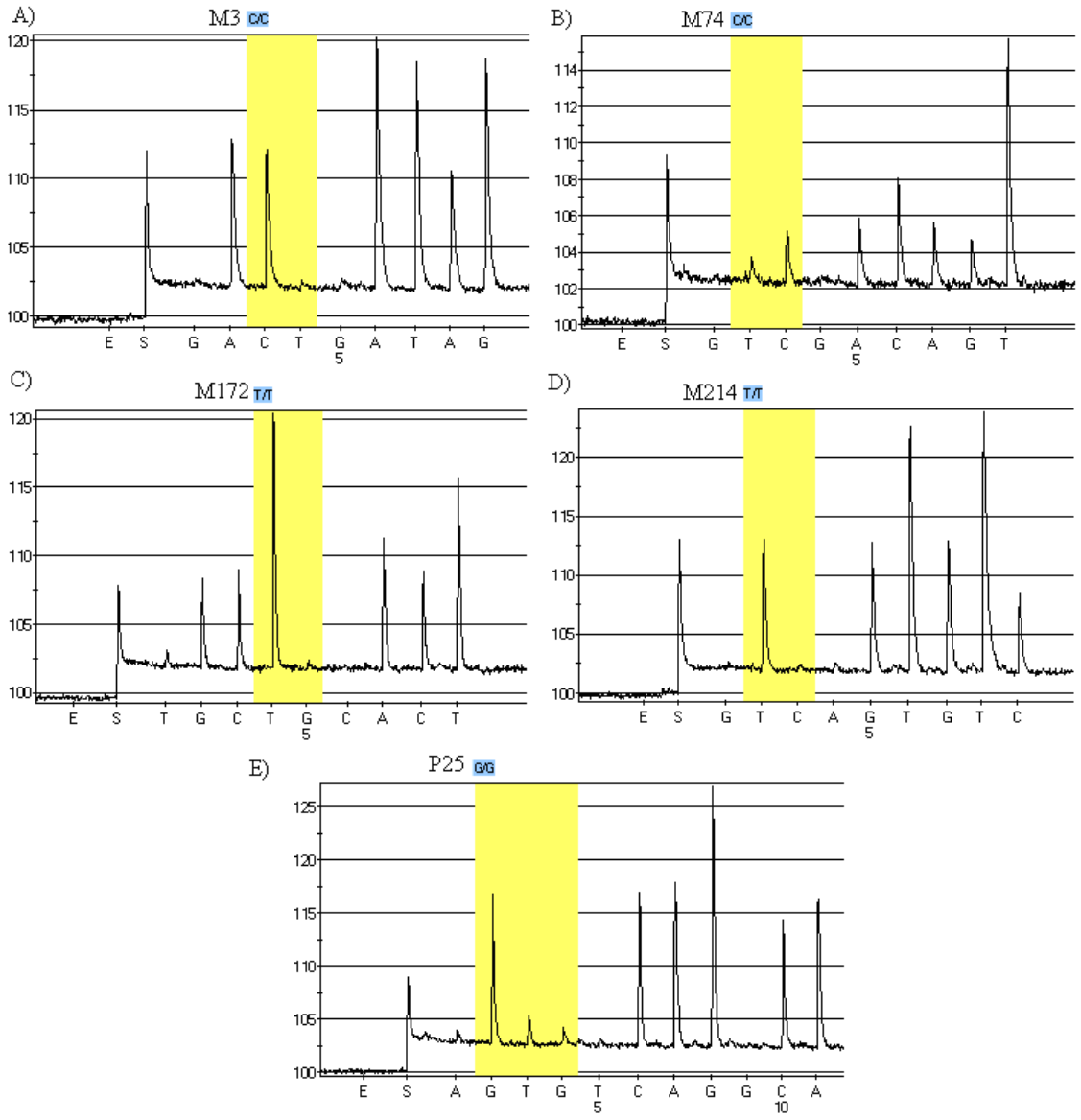


Figure 30. 10pg Male with 300ng female at M3, M74, M172, M207, and P25. Male individual B1 (10pg) with female individual 9947a (300ng) amplified with 13-plex outer primer set, and then analyzed using M3, M74, M172, M214, and P25 simplex amplification and pyrosequencing. A) Pyrogram produced for M3 showing unambiguous allele calling of the C allele. B) Pyrogram produced for M74 showing unambiguous allele calling of the C allele. C) Pyrogram produced for M170 showing unambiguous allele calling of the TTT homopolymer. D) Pyrogram produced for M214 showing unambiguous allele calling of the T allele. E) Pyrogram produced for P25 showing unambiguous allele calling of the GGG homopolymer.

CHAPTER 4. CONCLUSIONS

The goal was to evaluate Pyrosequencing as a technique for the analysis of Y-chromosome SNPs for forensic applications. Using Pyrosequencing to develop a system for analyzing multiple Y-chromosome SNPs from multiple sources, with minimal starting DNA (5pg) that can be analyzed as part of a mixture of male DNA in an overwhelming amount of female DNA (10pg of male DNA required in a 1 to 30,000 mixture). The system utilizes a nested PCR approach to amplify 13 Y-chromosome SNP loci, approximately 300 – 500 base pairs in size in the first round of amplification. The second round of amplification takes one to three of the starting 13 loci and amplifies them using the biotinylated primers necessary for Pyrosequencing (amplicons approximately 100 - 300 base pairs in size). Currently, four nested reactions are used to get maximum diversity from the 13-plex products: PMP1, PMP2, PMP3, and M74. The result is a system for the analysis of multiple Y-chromosome SNPs from multiple sources, using minimal starting DNA. Even in the presence of overwhelming amounts of female DNA, this system producing unambiguous sequenced Y-chromosome SNP haplotypes that can be used to distinguish a number of individuals from others who are unrelated.

All of the data produced from pyrosequencing indicated an ability to be used in forensic applications. Further work is required to determine if pyrosequencing is ideal for handling forensic applications, since multiple males sources, sources other than human, and degraded or damaged DNA was not examined. It is very likely that pyrosequencing will not have problems with multiple male sources, since it can currently provide the percentage of the peaks heights

observed in the allele calling region in reference to the other peaks observed in the sequence. DNA sources other than human are not likely to be a problem since such care was taken to ensure that all the primers designed for this system matched only to DNA sequences found on the human Y-chromosome and no other sequences found within the NCBI DNA sequence database. In terms of degraded DNA, it is more likely that the amplification procedures, most likely the PCR product size, would need to be modified to accommodate for this forensic obstacle. Some of the current outer PCR products are fairly large, in terms of forensic applications (339 – 485bp) since the nested PCR approach was adopted after many of the inner biotinylated primer pairs were already designed and evaluated, but could easily be redesigned to produce smaller PCR products.

LIST OF REFERENCES

1. *db SNP : Single Nucleotide Polymophism*, NCBI.

2. Tilford, C.A., et al., *A physical map of the human Y chromosome.* Nature, 2001. **409**(6822): p. 943-5.

3. Bachtrog, D. and B. Charlesworth, *Towards a complete sequence of the human Y chromosome.* Genome Biol, 2001. **2**(5): p. REVIEWS1016.

4. Cinnioglu, C., et al., *Excavating Y-chromosome haplotype strata in Anatolia.* Hum Genet, 2004. **114**(2): p. 127-48.

5. Underhill, P.A., et al., *The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations.* Ann Hum Genet, 2001. **65**(Pt 1): p. 43-62.

6. Johnson, C.L., et al., *Validation and uses of a Y-chromosome STR 10-plex for forensic and paternity laboratories.* J Forensic Sci, 2003. **48**(6): p. 1260-8.

7. Hall, A. and J. Ballantyne, *The development of an 18-locus Y-STR system for forensic casework.* Anal Bioanal Chem, 2003. **376**(8): p. 1234-46.

8. Sachidanandam, R., et al., *A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms.* Nature, 2001. **409**(6822): p. 928-33.

9. Giffard, P.M., et al., *Comparison of competitively primed and conventional allele-specific nucleic acid amplification.* Anal Biochem, 2001. **292**(2): p. 207-15.

10. Biosystems, P., *ABI Prism SNaPshot ddNTP Primer Extension Kit - Protocol*. 1999.

11. *SNP Genotyping*, University of Chicago - Department of Psychiatry Genotyping Shared Facility.

12. Ronaghi, M., *Pyrosequencing sheds light on DNA sequencing.* Genome Res, 2001. **11**(1): p. 3-11.

13. Pyrosequencing, A., *The Principle of Pyrosequencing*. 2000.

14. Comey C, P.K., Smerick J, Sobieralski C, Stanley D, Baechtel F, *Extraction Strategies for Amplified Fragment Length Polymorphism Analysis.* Journal of Forensic Science, 1994. **39**: p. 1254-1269.

15. Skaletsky, S.R.a.H.J., *Primer3 on the WWW for general users and for biologist programmers.* Bioinformatics Methods and Protocols: Methods in Molecular Biology, 2000: p. 365-386.

16. Pyrosequencing, A., *Technical Support*. http://techsupport.pyrosequencing.com/v2/index.asp.

17. Pyrosequencing, A., *Sample Preparation Guidelines for the PSQ™96 and PSQ 96MA Systems*. 2003, Pyrosequencing, AB: http://techsupport.pyrosequencing.com. p. 24.

18. *A nomenclature system for the tree of human Y-chromosomal binary haplogroups.* Genome Res, 2002. **12**(2): p. 339-48.

19. Underhill, P.A., et al., *Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography.* Genome Res, 1997. **7**(10): p. 996-1005.

20. Underhill, P.A., et al., *Y chromosome sequence variation and the history of human populations.* Nat Genet, 2000. **26**(3): p. 358-61.

21. Hammer, M.F., et al., *Hierarchical patterns of global human Y-chromosome diversity.* Mol Biol Evol, 2001. **18**(7): p. 1189-203.

22. Su, B., et al., *Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age.* Am J Hum Genet, 1999. **65**(6): p. 1718-24.

23. Underhill, P.A., et al., *A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 196-200.