CALIFORNIA STATE UNIVERSITY SAN MARCOS

THESIS SIGNATURE PAGE

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

IN

BIOTECHNOLOGY

THESIS TITLE: Validation of YFiler Plus Amplification Kit for the San Diego Police Department

AUTHOR: Alexander Phan

DATE OF SUCCESSFUL DEFENSE: July 27th, 2017

THE THESIS HAS BEEN ACCEPTED BY THE THESIS COMMITTEE IN
PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF
SCIENCE IN BIOTECHNOLOGY.

Betsy Read THESIS COMMITTEE CHAIR	Betsy Read SIGNATURE	08/03/2017 DATE
James Jancovich THESIS COMMITTEE MEMBER	 SIGNATURE	08/03/2017 DATE
Shawn Montpetit THESIS COMMITTEE MEMBER	 SIGNATURE	08.03.2017 DATE
Patrick O'Donnell THESIS COMMITTEE MEMBER	 SIGNATURE	8/7/17 DATE

Validation of the YFiler Plus Amplification Kit for the San Diego Police Department

San Diego Police Department – Forensic Biology
Alexander Phan
July 10th, 2017

Faculty Advisors
Project Chair: Dr. Betsy Read
Committee Member: Dr. James Jancovich
Committee Member: Dr. Patrick O'Donnell

Professional Science Masters
California State University, San Marcos

Table of Contents

List of Figures

Executive Summary

Title: Validation of the YFiler Plus Amplification Kit for the San Diego Police Department
Company Sponsor: San Diego Police Department – Forensic Biology
Author: Alexander Phan
Date: July 10th, 2017
Professional Science Masters
California State University, San Marcos

Currently, the San Diego Police Department Forensics Biology Unit utilizes the AmpFlSTR® YFiler PCR Amplification Kit from Applied Biosystems. However, they have been looking to upgrade to the new YFiler Plus Amplification Kit from Thermo Fisher Scientific because the current kit is outdated and YFiler Plus offers more discrimination power due to the increased number of markers and improved performance with degraded samples. Hence the two specific aims of this project were: 1) to perform a validation by conducting multiple studies to compare the efficacy of the new typing kit versus the current typing kit, 2) document and write reports of validation results from each study performed. The validation began with finding the optimal amplification conditions on the Veriti (Thermalcycler) and finding the best parameters for capillary electrophoresis on the 3500 Genetic Analyzer. Afterwards, the following studies were carried out: baseline, sensitivity, stutter, balance, concordance, stochastic, and DNA mixtures. Once the validation was finished, a summary was written for each study to elaborate on how and why and the experiments were carried out, discuss results and analysis, and present meaningful conclusions. Once the expectations were met, the new kit was implemented into forensic casework.

The forensic biology unit currently employs the AmpFlSTR® YFiler PCR Amplification Kit, a kit that was implemented and validated in May 2007. The forensic biology unit decided to validate the YFiler Plus Amplification Kit as a potential replacement as it promises to provide an improved baseline and higher recovery of alleles. The YFiler PCR Amplification Kit amplifies the same seventeen loci markers that are in the Yfiler Plus Kit (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385 a/b, DYS393, DYS391, DYS439, DYS635, DYS392, YGATAH4, DYS437, DYS438, and DYS448). Upon verifying the amplification of these markers, the ten additional Y-STR loci (DYS576, DYS627, DYS460, DYS518, DYS570, DYS449, DYS481, DYS387S1 a/b, and DYS533) included in the Yfiler Plus Kit will be tested. These new markers are important as they offer greater discrimination when searching for the male suspect, which increases the statistical probability of obtaining the correct identification.

Overall, the YFiler Plus Amplification Kit from Thermo Fisher Scientific is a 6-dye multiplex assay that allows for amplification of 27 Y-STR markers. It is created for forensic laboratories that carry out challenging sexual assault cases and it optimizes the analytical processes within forensic laboratory casework. With YFiler Plus Kit, the following features resulted: markers that mutate more frequently and amplify more efficiently, improved DNA results from degraded samples that contain PCR inhibitors, increased genotyping accuracy through the developed allelic ladder and virtual bins, and enhanced protocol efficiency. Even in the presence of an

abundance of female DNA, the new Yfiler Plus Kit readily detected male DNA from mixed samples.

Introduction

Suspects were able to get away with crimes and their actions had no consequences until a British forensic pathologist named Dr. Bernard Spilsbury solved the "Dr. Crippen Case" in 1910 (Jacobs, 2007). He solved this case by providing expert testimony and showing detailed microscopic study of the suspect's scar (Jacobs, 2007). As a result, Dr. Bernard Spilsbury is deemed the father of forensics because he applied scientific analysis and began the era of application of DNA science.

Imagine a sexual assault case inside a residential home where the police arrive and the male suspect has fled. The police begin their investigation and examine the house and the environment for evidence collection. This is where forensic biology plays a role during the crime scene investigation. Forensic biology is the analysis of multiple biological characteristics among individuals that are found during evidence collection at the scene of the crime (Lewontin, 1991). With modern-day DNA technology, forensic scientists have the capability to obtain information about a person with minimal amounts of DNA.

Historically, sexual assault cases were difficult to solve in the absence of a single method to identify a suspect. In 1985, an English geneticist named Alec Jeffreys was the first person to characterize DNA typing or DNA profiling (Butler, 2005). He noted that there were specific regions of DNA that have unique sequences repeated multiple times. In addition, he was able to provide distinctive evidence for a specific suspect from a list of individuals by examining a number of repeated sections within a DNA sample. These repeated sections of a DNA sample became known as variable number of tandem repeats. Dr. Jeffreys created the first set of human identity tests, which are now widely used in forensic applications and law enforcement.

One special forensic technique is Y-STR (Y-chromosome short tandem repeat) testing, which provides male genetic profiles after analyzing evidence collected from a rape crime scene (Bailey-Van Houten, 2005). For Y-STR testing, DNA can be collected from sperm cells, semen, and/or saliva. In order to carry out the Y-STR test, DNA is first extracted and purified and qPCR (quantified polymerase chain reaction) is used to determine the total amount of DNA present. After target sequences are amplified via end-point PCR (polymerase chain reaction), products are sized using a capillary electrophoresis system. Finally, an electropherogram is developed to showcase the DNA profile that presents a unique set of male alleles.

For the SDPD forensic biology unit to implement the YFiler Plus Amplification Kit, it must adhere to the FBI quality assurance standards and interpretation guidelines for Y-Chromosome STR testing. Standards were set forth by an international and national of forensic scientists (Scientific Working Group on DNA Analysis Methods, SWGDAM) convened and charged with establishing guidelines and standards for crime laboratories. SWGDAM is also responsible for delivering recommendations and revisions for the FBI quality assurance standards (FBI, 2011). Overall, implementations of new DNA technology or for prospective crime laboratories to establish themselves, it is required by federal law to follow the FBI's quality assurance

standards. In addition, crime labs must adhere to section 8 of the FBI's quality assurance standards, which lays out details on how to conduct internal validations for DNA analysis (FBI, 2011).

The SDPD's current DNA amplification kit (YFiler PCR) provides sufficient scientific results for sexual assault cases during casework. However, the current kit is outdated and is not as effective when working with degraded or dirty samples. The new kit provides multiple distinct loci for analysis, which allows for greater discrimination when searching for a male suspect. It also incorporates up-to-date PCR technology and has the capability to function optimally with degraded samples or contaminated samples that contain PCR inhibitors. To test the capabilities of the YFiler Plus Amplification kit, an internal validation was conducted. The kit demonstrated the capacity for improved discrimination, easier workflow efficiency for faster result times, and improved peak height ratios during GeneMapper ID-X (v1.4) analysis. Once these have been met, the objective of the internal validation study was accomplished and the YFiler Plus Amplification Kit can be incorporated into the SDPD forensic biology unit's technical protocol.

Materials and Methods

Single source (reference) samples were primarily used during the internal validation of this study. Following the manufacturer's protocol on preparing the amplification reactions, the total volume of the YFiler Plus Master Mix (10.0 µL per reaction) and YFiler Plus Primer Set (5.0 µL per reaction) was determined based on the number of reactions to be performed. For example, if there were 10 reactions to be prepared, then a total of 100 µL of YFiler Plus Master Mix and a total of 50 µL of YFiler Plus Primer Set will be combined. A total 15 µL of the reaction mixture (Master Mix & Primer Set) were distributed per number of reactions into the respective wells of the MicroAmp optical 96-well reaction plate. For the negative control, 10 µL of TE buffer were added into one of the wells. The targeted amount of 0.6 ng/µL of DNA was used in each reaction for a final volume of 10 µL (DNA & TE Buffer). The designated amount of 007 control DNA was added to act as the positive control. The final reaction volume was 25 µL. The plate was covered with strip caps to prevent evaporation. The plate was centrifuged at 3000 rpm for approximately 20 seconds to collet reaction components, before being placed in the Applied Biosystems' Veriti 96 Well Thermal Cycler by Applied Biosystems. Cycling parameters included an enzyme activation step at 95°C for 1 minute followed by 30 cycles of denaturation at 94°C for 4 seconds, annealing at 61.5°C for 1 minute and elongation at 60°C for 22 minutes.

Amplification products were sized using capillary electrophoresis. Following the manufacturer's protocol, the volumes of Formamide (9.6 µL per reaction) and GeneScan 600 LIZ Size Standard v2.0 (0.4 µL per reaction) was determined based on the total number of reactions. For example, if there were 10 planned reactions, then 96 µL of Formamide and 4 µL of GeneScan Size Standard will be combined. A total of 10 µL of the reaction mix (Formamide/GeneScan Size Standard mix) and 1 µL of PCR product were combined and pipetted per number of reactions into each respective well of the MicroAmp Optical 96-well reaction plate. Also, one of the respective wells contained 10 µL of the reaction mix and 1 µL of the allelic ladder. The plate was sealed with a septa and centrifuged. The plate was heated for 3 minutes at 95°C. Then the

reaction plate was snap cooled on an ice block for 3 minutes. The reaction plate was placed into the 3500 Genetic Analyzer to undergo capillary electrophoresis.

The initial part of the validation began with finding optimal conditions for both amplification (cycle number) on the Veriti and capillary electrophoresis (injection times & voltages) on the 3500 Genetic Analyzer. This study was carried out using approved SDPD employee reference samples and blank samples. Two male-approved validation samples were amplified using the following DNA template amounts: 0.1 ng, 0.25 ng, 0.5 ng, 1.0 ng, 1.5 ng. These two sets of DNA template amounts were run initially at 27 cycles using the Veriti thermal cycler. Then another set of the same DNA template amounts were run at 28 cycles. The same was done for 29 cycles. Lastly, one of the two sets of DNA template amounts were run under 30 cycles. Another run with the same validation samples at the same DNA template were carried out on the current YFiler PCR kit at 30 cycles for comparison purposes. The amplification products from the DNA samples were injected for capillary electrophoresis on the 3500 Genetic Analyzer.

However, there was a change to the parameters for the 3500 instruments in order to optimize the data. One set of amplified DNA samples that were run for 29 cycles were sized on the 3500 capillary electrophoresis instrument at: 1) 1.2 kilovolts for 15 seconds, 2) 1.2 kilovolts for 16 seconds, 3) 1.5 kilovolts for 16 seconds, 4) 1.5 kilovolts for 20 seconds, and 5) 1.5 kilovolts for 24 seconds. A second set of amplified DNA samples were run for 30 cycles and sized using the 3500 instrument at: 1) 1.5 kilovolts for 16 seconds, and 2) 1.5 kilovolts for 10 seconds. DNA peaks and PCR artifacts obtained from the different runs were assessed.

A baseline study was carried out to determine the analytical threshold or the limit of detection of YFiler Plus Amplification Kit data. The analytical threshold is the minimal RFU height requirement required to distinguish signal from background noise. The baseline study was evaluated with approved SDPD employee reference samples and blank samples. Positive controls, negative controls, and 62 DNA-containing samples were amplified using the amount of input DNA and cycling parameters determined from the optimal conditions study described above. The recommended parameters were: 10 µL of TE for blanks, 0.6 ng of DNA and 30 cycles. The amplification product from the positive controls, negative controls, and 62 DNA-containing samples were injected for capillary electrophoresis on both 3500 Genetic Analyzers at 1.5 kilovolts for 10 seconds. During GeneMapper ID-X v1.4 analysis, the analytical threshold of 1 RFU was applied to all samples and blanks. Immediately thereafter, these datasets were exported from GeneMapper ID-X (v1.4) and imported into STR-Validator software program. The STR-Validator software was utilized to quickly calculate the analytical threshold. In addition, the software allowed for removal of DNA peaks, PCR artifacts (stutter peaks, minus-A peaks), and spectral artifacts. After the analytical threshold was calculated, the data was further evaluated to confirm the reliability of the analytical threshold. This was done by searching for the top 10 highest noise peaks in the overall dataset. It allowed for detecting the highest signal limit of background noise, so that the analytical threshold can be established above this limit.

A sensitivity study was conducted to determine the level where allelic dropout (a phenomenon that occurs due to low level of DNA template where the allele fails to be detected or loss of amplicon product) is noted and that which results in an increase in amplicon artifacts and/or the saturation point for signal. Plus, the optimal level for the designated DNA template range was

found. This study established a general range of DNA concentrations that confirms the reliability of DNA typing results when using the YFiler Plus Amplification kit in conjunction with the 3500 Genetic Analyzer. Ten male reference samples were amplified at various DNA input levels for 30 cycles with the YFiler Plus kit. In addition, the two male reference samples from the optimum cycle number and capillary electrophoresis parameters study was also examined for this study. These ten samples were amplified at the following DNA template ranges: 60 picograms (pg), 80 pg, 100 pg, 150 pg, 200 pg, 300 pg, 400 pg, and 800 pg. Plus, these same ten samples were run again between 12 pg to 20 pg in order to obtain results with increased allelic dropouts. The other two male reference samples were amplified at 100 pg, 250 pg, 500 pg, 1 ng, and 1.5 ng. These DNA template amounts were chosen to determine ideal input levels, dropout levels, and saturation levels. After amplification, these samples were injected for capillary electrophoresis on both 3500 Genetic Analyzers at 1.5 kilovolts for 10 seconds. A detection threshold of 100 RFU was applied during GeneMapper ID-X v1.4 analysis. During the analysis process, peak heights and artifacts were evaluated in order to determine the ideal DNA target amount.

A stutter study was carried out to examine the types of DNA artifacts and various levels of stutter peaks that are observed using the YFiler Plus Amplification Kit on the 3500 Genetic Analyzers. Stutter peak products are an amplification phenomenon that is found 4 bases shorter than the corresponding main allele peak (Butler, 2005). Positive controls, negative controls, and 62 male reference samples were amplified using 0.6 ng of DNA template for 30 cycles. Then capillary electrophoresis was carried out at the same parameters as those used for analytical threshold study on both 3500 Genetic Analyzers. The data produced was analyzed in GeneMapper ID-X (v1.4) using a 100 RFU detection threshold. Data from the GMID-X (v1.4) was imported into STR-Validator to calculate stutter percentage across all the markers.

The goal of the balance study was to evaluate the inter- and intra-locus dye balance by analyzing the resulting peak heights following amplification with the YFiler Plus Amplification kit. Amplification and capillary electrophoresis was carried out as previously described in the stutter study. The resulting data was further examined in GeneMapper ID-X (v1.4) using an analytical threshold of 100 RFU. The inter- and the intra-locus dye balance was determined by calculating the average peak heights across all the loci.

For the concordance study, the objective was to confirm that the DNA typing results obtained with the YFiler Plus Amplification Kit are consistent with the DNA typing results obtained with AmpFlSTR® YFiler PCR Amplification Kit. The study was also performed to verify results obtained using the new kit for analysis of NIST (National Institute of Standards and Technology) traceable standards match the known published results. Sixty male reference samples and two NIST samples (Components B and C) were amplified on the Veriti thermalcycler for 30 cycles with a target input of 0.6 ng of DNA using the YFiler Plus Amplification Kit. Capillary electrophoresis and data analysis was carried out as previously described in the balance study. The YFiler Plus data was compared with the 17 loci shared with AmpFlSTR® YFiler PCR.

The stochastic study was carried out to determine the probability levels at which drop-out of genotypes occur and to calculate the threshold after single source samples are amplified with the YFiler Plus Amplification Kit. The stochastic threshold corresponds to the peak is above this threshold, then the sister allele will be present. Amplification was carried out as described in the

stutter study. In addition, a dilution series of 10 samples were amplified at a range of target amounts from ~12 pg to 800 pg. Capillary electrophoresis was performed on the amplified DNA products using both 3500 Genetic Analyzers as previously described in the stutter study. The overall stochastic threshold for YFiler Plus was calculated using James Curran's online software program. In addition, the stochastic threshold was determined for each color set (Blue, green, yellow, red, purple). This was done by importing data from GMID-X (v1.4) into the online software. Using the software settings shown below, the overall stochastic threshold was generated. In addition, the stochastic threshold was calculated for each color dye (blue, green, yellow, red, and purple).

Lastly, a DNA mixture study was performed to assess the male-to-male, male-to-female, and three male mixtures combined in various ratios. In addition, the male-to-female mixture samples were further analyzed to determine the ratio whereby female DNA inhibits a male DNA sample. Positive controls, negative controls, and a total of ten male employee reference samples were utilized. Male-to-male mixtures were created using the following ratios: 6:1, 4:1, 3:1, and 2:1. These samples were amplified in duplicate with a total of 600 pg, 300 pg, 200 pg, 150 pg, and 100 pg of DNA. Male-to-female mixtures were created using a constant amount of male DNA (600 pg and 100 pg) with increasing amounts of female DNA to target ratios of: 1:1, 1:10, 1:100, 1:1000, and 1:10000. Three male mixtures were created in the following ratios: 3:2:1, 3:1:1, and 2:1:1. These three male mixtures were amplified eight times each at 300 pg, 200 pg, 150 pg, and 100 pg. All of the mixtures (male-to-male, male-to-female, and three males) were amplified using YFiler Plus Amplification kit. Capillary electrophoresis was performed on both 3500 Genetic Analyzers following the thermal cycling conditions and capillary electrophoresis conditions provided from the initial study. Data analysis was carried out using the GeneMapper ID-X (v1.4) software with the baseline determined from the analytical threshold study.

Results

Sample Name	# of Artifacts Found (29	Average Peak Height (29 Cycles)	Number of Allelic Dropouts	Sum of Peak Heights (29 Cycles)
007_Control_1.0	76	20473	0	470884
020v(Staff)_100_Tube I	0	313	2	7815
020v(Staff)_250_Tube I	5	939	0	25343
020v(Staff)_500_Tube J	19	1930	0	52114
020v(Staff)_1000_Tube K	22	3342	0	90236
020v(Staff)_1500_Tube L	33	13412	0	375541
373v(Staff)_100_Tube I	0	432	0	11668
373v(Staff)_250_Tube I	3	641	0	17298
373v(Staff)_500_Tube J	24	2579	0	69620
373v(Staff)_1000_Tube K	25	3346	0	90343
373v(Staff)_1500_Tube L	30	11855	0	308233
007_Control_2.5	21	2008	0	54223
007_Control_5.0	N/A	N/A	N/A	N/A
007_Control_10.0	45	9911	0	257690

Sample Name	# of Artifacts Found (Current Kit)	Average Peak Height (Current Kit)	Number of Allelic Dropouts	Sum of Peak Heights (Current Kit)
007_Control_1.0	1	2070	0	35191
020v(Staff)_100_Tube I	0	801	0	13621
020v(Staff)_250_Tube I	5	4750	0	80750
020v(Staff)_500_Tube J	7	5278	0	89733
020v(Staff)_1000_Tube K	9	8982	0	152687
020v(Staff)_1500_Tube L	62	19347	0	251507
373v(Staff)_100_Tube I	1	1747	0	27949
373v(Staff)_250_Tube I	5	4411	0	74994
373v(Staff)_500_Tube J	9	6337	0	107724
373v(Staff)_1000_Tube K	11	9136	0	146180
373v(Staff)_1500_Tube L	33	28435	0	398092
007_Control_2.5	5	6156	0	98502
007_Control_5.0	9	10126	0	162014

Sample Name	# of Artifacts Found (1.5kv at 16)	Average Peak Height (1.5kv at 16)	Number of Allelic Dropouts (1.5kv at 16)	Sum of Peak Heights (1.5kv at 16)	Peak Height Increase from 29 cycles to 30 cycles at 1.5kV for 16 secs	Average Peak Height Increase from 29 cycles to 30 cycles at 1.5kV for 16 secs
007_Control_1.0	0	687	0	18555	1.11526	1.9581
373v(Staff)_100_Tube H	0	896	0	50163	1.68105	
373v(Staff)_250_Tube I	0	1214	0	31567	0.9712	
373v(Staff)_500_Tube J	0	4843	0	130748	2.35554	
373v(Staff)_1000_Tube K	9	12217	0	305424	4.07098	
373v(Staff)_1500_Tube L	21	18780	0	450712	1.81134	
007_Control_2.5	2	2718	0	76113	1.83897	
007_Control_5.0	4	5015	0	135405	1.97286	
007_Control_10.0	10	12277	0	319200	1.80571	

Figure 1 – First table listing observations for 29 cycles (YFiler Plus) and 30 cycles (YFiler PCR) and the second table listing observations for 30 cycles (YFiler Plus)

To identify optimal conditions for amplification and capillary electrophoresis, the two male samples were run for 27, 28, and 29 (Figures 1) cycles separately using identical amounts of DNA template (100, 250, 500, 1000, and 1500 pg). When all template DNA was amplified using 27 and 28 cycles, a significant number of loci did not amplify. There were a low number of artifacts on all parameters. Artifacts are by-products of the PCR process and resemble DNA peaks. However, these artifact peaks need to minimized as much as possible. The largest peaks were noted when using 29 cycles, however, they did not yield the same sensitivity as the previous kit, YFiler PCR. After changing the 3500 Genetic Analyzer parameters for capillary electrophoresis to optimize the RFU values of the DNA peaks, the data showed a general increase in RFU values from 1.2 kV at 15 seconds to 1.2 kV at 16 seconds to 1.5 kV at 16 seconds (Figure 2). The average peak height increase from 15 seconds to 16 seconds at 1.2 kV was calculated to be 1.33x bigger (Figure 2). The average peak height increase from 1.2 kV to 1.5 kV at 16 seconds was calculated to be 1.09x bigger (Figure 2). These peak heights were still not consistent with the previous kit, so additional 3500 Genetic Analyzer parameters were explored. The average peak height increase from 20 seconds to 24 seconds at 1.5kV was calculated to be 1.03x bigger (Figure 3). The RFU values were still not in the sensitivity range as the previous kit; therefore, the samples were run at 30 cycles. The average peak height increase from 29 cycles to 30 cycles at 1.5 kV at 16 seconds was calculated to be 1.95x bigger (Figure 4), which fell within expectations. After analyzing the data, 30 cycles showed the same sensitivity as the previous kit at 1.5 kilovolts at 16 seconds and 1.5 kilovolts at 10 seconds. By increasing cycle numbers,

injection times, and injection voltages, there was a decrease in allelic dropouts. However, there was an increase in detected artifacts at the 1.0 ng and 1.5 ng DNA template amounts across all the amplification cycles. When comparing the YFiler Plus results to the current YFiler kit results for the same DNA samples at the same DNA inputs, there were more artifacts found with the AmpFlSTR® YFiler PCR kit amplified products compared to the YFiler Plus kit amplified products. Based on the results of this study, it was determined that the optimal number of cycles is 30 and the designated 3500 capillary electrophoresis parameters will be 1.5 kilovolts at 10 seconds.

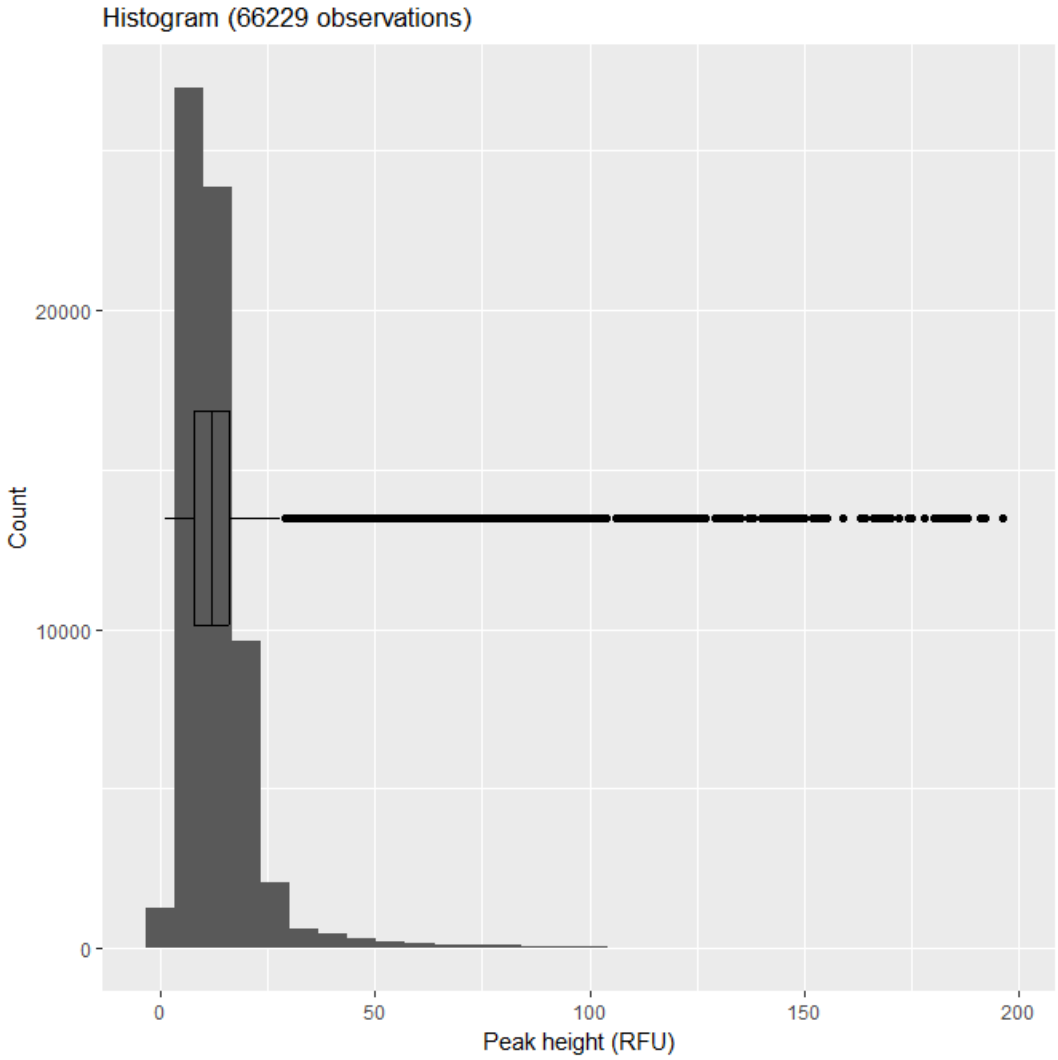

Figure 2 – STR-Validator's histogram showing the number of noise peaks found at various peak heights from the 62 male reference samples. A significant number of noise peaks fell below 50 RFU,

Global.AT2	Global.B.AT2	Global.G.AT2	Global.Y.AT2	Global.R.AT2	Global.P.AT2
64	53	71	79	59	61

Global.Mean	Global.Sd
13.58432	11.33595

$$AT_{M2} = r, when \frac{n_{s \leq r}}{N} > t$$

Figure 3 - STR-Validator's calculated analytical threshold values and global mean and global standard deviation for all 62 male samples. The 1st column displays the overall global analytical threshold values and the 2nd-6th columns displays the global analytical per dye channel: blue, green, yellow, red, and purple respectively.

A histogram was generated from STR-Validator to showcase the noise distribution (Figure 5). The top 10 noise peaks did not exceed the analytical threshold indicating they represent true noise. Therefore, the calculated analytical threshold was confirmed to be reliable (Figure 7). Noise peaks are background peaks and are not called during GeneMapper ID-X analysis. STR-Validator also utilizes 4 analytical threshold formulas to designate the RFU levels that exceed noise (Figure 6) and represent amplicon signal. However, only of 1 of the analytical threshold formulas will be looked at. It produced an average signal RFU values of 13.58 +/- 11.34. Based upon the data collected from this study, 100 RFU was operationally defined as the analytical threshold for DNA casework due to pragmatic reasons. There needs to be a balance between the background noise peaks and real DNA peaks. Also, future possible spectral issues or future higher background noise peaks have to be accommodated

The sensitivity of electropherograms were examined for all the samples to find the peak heights, allelic dropouts, artifacts, and saturation points. Saturation was determined to be 32,000 RFU for the 3500 Genetic Analyzers. The detection limit for full DNA profiles is approximately 60-100 pg. Spectral artifacts were increasingly observed when peak heights exceeded 10,000 RFU and generally occurred in samples that contained above 800 pg of DNA. The optimal range for robust DNA profiles is between 400-800 pg (Figure 8) with the recommended amount of DNA input being approximately 600 pg of DNA.

Five different types of stutter were found in the DNA profiles amplified with YFiler Plus. Stutter is a recurring phenomenon based on the number of base pairs less than the parent. Within this dataset, there were many n- 1 and n+1 repeat stutters while there were the occasional n-0.5, n-2, n+2 repeat stutters. When using GMID-X, the default stutter filters were used to examine the occurrences across all the samples. The default GMID-X filter will be adjusted accordingly, so that there will be a balance between filtering the majority of stutter peaks while preserving allelic peaks in a mixture where DNA contribution can be unbalanced. Stutter patterns differ from allele to allele. This validation only contains 62 samples; therefore, it does not contain all the possible combinations of alleles that could be observed in casework. Results suggest larger alleles typically display higher stutter peaks.

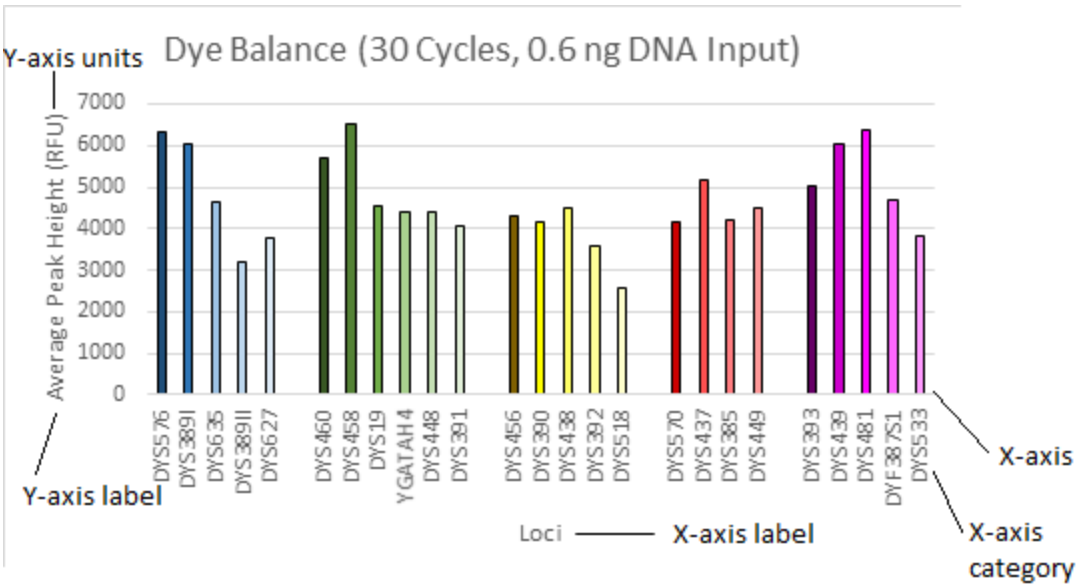

Figure 4 – A bar graph displaying the dye balance between the loci

To define comparisons between the different dye colors and each individual locus in the balance study, the average peak height of each loci was calculated across all samples. In general, perfect balance is when all the peak heights are all the same RFU value across all the loci or when any locus peak height divided by another locus peak height is 1. Figure 4 helped determine expectations for behavior of the kit under the parameters established from the initial study and gave a representation on how the different loci amplify within the kit. In addition, it will help develop interpretation guidelines in the forensic biology unit.

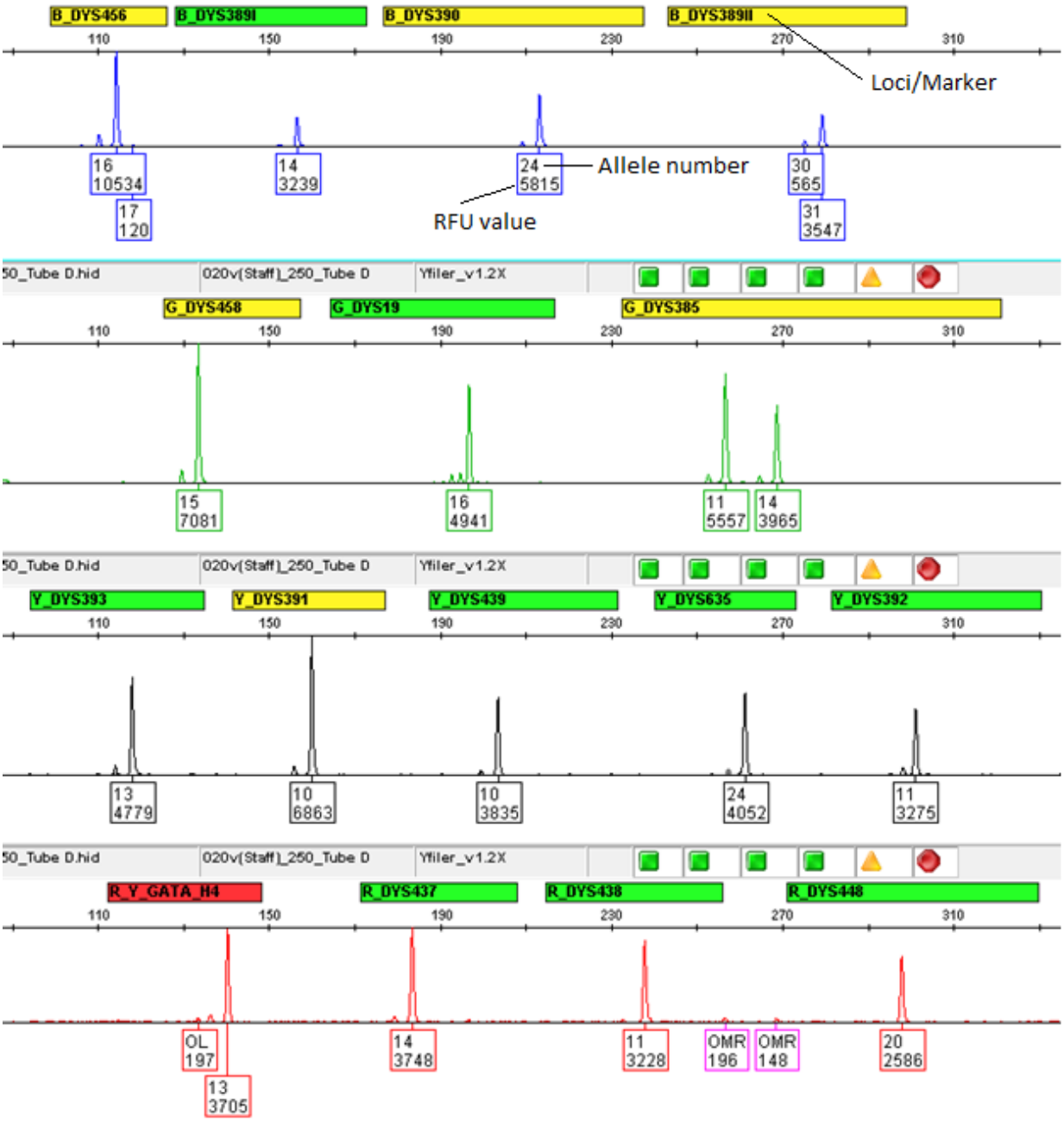

Figure 5 – Male reference employee DNA profile (YFiler PCR 30 Cycles) from GeneMapper ID-X (v1.4) for concordance study. Each peak is distinguished by color. All the peaks that are colored blue are within the blue dye, which are fluorescently tagged blue on the primers when it attaches to the specified loci. All the peaks that are colored green are within the green dye, which are fluorescently tagged green on the primers when it attaches to the specified loci. All the peaks that are colored black are within the yellow loci, which are fluorescently tagged yellow on the primers when it attaches to the specified loci. All the peaks that are colored red are within the red loci, which are fluorescently tagged red on the primers when it attaches to the specified loci.

Figure 6 – Male employee reference DNA profile (YFiler Plus 29 Cycles) from GeneMapper ID-X (v1.4) for concordance study. Each peak is distinguished by color. All the peaks that are colored blue are within the blue dye. All the peaks that are colored green are within the green dye. All the peaks that are colored black are within the yellow loci. All the peaks that are colored red are within the red loci. All the peaks that are colored purple are within the purple loci, which are fluorescently tagged purple on the primers when it attaches to the specified loci. The purple loci are a new additional feature of YFiler Plus compared to YFiler PCR.

The genotypes for the NIST standard reference material (Components B and C) were 100% concordant at all loci after being amplified with the YFiler Plus kit. The genotypes obtained following YFiler Plus amplification of the 60 male reference samples were 100% concordant at the markers previously analyzed with AmpFlSTR® YFiler kit (Figures 12 & 13). The positive control (DNA control 007) was consistent with all the profiles in every amplification performed with the YFiler Plus kit. There are ten new additional Y-STR markers (DYS576, DYS627, DYS460, DYS518, DYS570, DYS449, DYS481, DYF387S1 a/b, and DYS533) within the YFiler Plus kit. These markers were not able to undergo testing for concordance with the AmpFlSTR® YFiler PCR kit. However, the 62 male samples were amplified repeatedly and results demonstrated coherence with the YFiler Plus kit. All male sample profiles were shown to have successful DNA typing at all 27 markers. DNA typing results acquired using YFiler Plus is

100% concordant with the results acquired using AmpFlSTR® YFiler PCR. This verifies the expected results and provides additional genotype characterization due to the addition of ten new Y-STR markers.

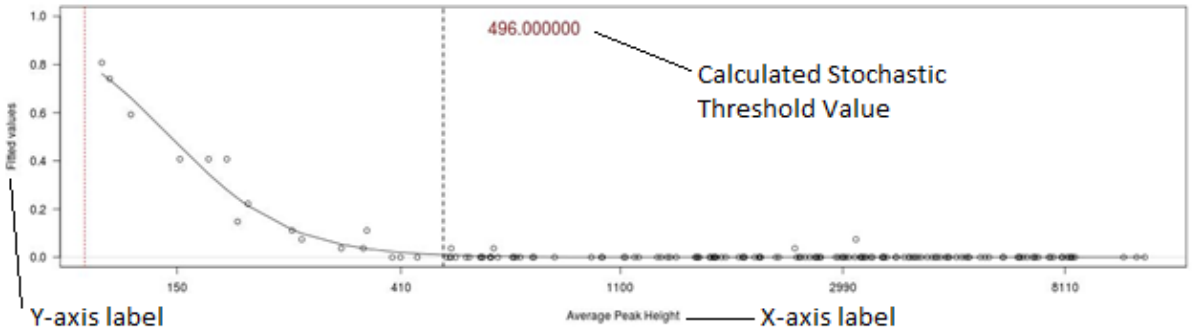

Figure 7 – The calculated overall stochastic threshold value from James Curran's software program for the all the loci combined. James Curran's software program creates a regression curve to the data. With a given alpha (0.99 for the graph above), it can be used to calculate the probability of drop out. The circles represent the datapoints from the 62 reference samples and the dilution series samples. The x-axis represents the average peak heights of the data points. The y-axis represents the probability of drop-out. The dotted line is the stochastic threshold, which is 496 RFU.

Complete haplotypes in single source samples was observed at and above 100 pg of DNA. Allelic dropout was observed ate template levels below 100 pg of DNA input. Overall, the stochastic threshold for YFiler Plus was calculated to be 496 RFU (Figure 14). This fell within expectations and does not require further adjustments. At 496 RFU and above, there is an exceedingly small chance that drop out has occurred. It was also important to calculate the stochastic threshold values for the blue, green, yellow, red, and purple dyes (Figures 15-19). Blue was 552 RFU. Green was 463 RFU. Yellow was 506 RFU. Red was 376 RFU. Purple was 478 RFU. Occurrence of dropout begins to take place below 100 pg of DNA input and the dropout rate increases as DNA input decreases. Also, probability of drop out increases below 496 RFU.

Figure 8 – Male-Female 1:10000 DNA profile sample from GeneMapper ID-X (v1.4)

After running mixture samples with YFiler Plus, all of them were observed for its performance according to the created ratios. The 2 male mixtures and 3 male mixtures was assessed and performed to expectations from the created mixturre ratios. The male-female mixtures (1:1, 1:10, 1:100, and 1:1000) samples had full, robust male DNA profiles without allelic dropouts. Only 50% of the male DNA profile was detected, however, at a male-female ratio of the 1:10000 (Figure 8). Therefore, female DNA acts as an inhibitor above the 1:1000 level for male-female DNA ratios. Overall, the results of this mixture study provide a large range of contributor ratios and input amount that can be used to evaluate its capabilities when using STRmix.

Discussions and Conclusions

For future validation, developing mixture interpretation guidelines for YFiler Plus would be the next step. SDPD's forensic biology unit wishes to utilize Oregon State Police's YFiler Plus mixture interpretation guidelines. In order to carry this out, the YFiler Plus mixture interpretation guidelines from Oregon State Police must be used on the mixture samples from this validation. If the mixture samples meet the guideline's expectations, then SDPD's forensic biology unit can incorporate the same guidelines. If the mixtures samples do not meet the guideline's expectations, then SDPD's forensic biology unit must adjust the guidelines accordingly. Once the mixture interpretation guidelines have been completed, SDPD's forensic biology unit can add the YFiler Plus Amplification Kit to their technical manual for DNA casework.

Lastly, the updated CalDOJ Y-Mixture tool would be explored for potential capabilities. It is an excel database tool used for both single source and mixed Y-STR haplotypes to be searched against the US Y-STR haplotype database to obtain estimates of the frequency of occurrence for the haplotypes.

After examining capabilities and limits of YFiler Plus through each study conducted, the kit was incorporated into the forensic biology unit's technical manual. With this kit in place, there is increased discrimination for DNA identification due to the increased number of loci. YFiler Plus was shown to function optimally in mixture samples that contained a large amount of female DNA, which resulted in full male DNA profiles. Amplification time is much faster with YFiler Plus than YFiler PCR, which enhances protocol efficiency. Ultimately, the YFiler Plus Amplification Kit from Applied Biosystems optimizes the analytical DNA process for forensic casework.

References

Jacobs W, De Leeuw M. **Forensics, History**. *JAMA.* 2007;297(2):215-220. doi:10.1001/jama.297.2.216

Balding, David J. 2005. **Statistics in Practice: Weight-of-Evidence for Forensic DNA Profiles.**

Bailey-Van Houten, Lora. **Evaluation of Y-chromosome short tandem repeat multiplexes for forensic DNA analysis.** California State University, Fresno, ProQuest Dissertations Publishing, 2005. 1428628.

Lewontin, R.C., Hartl, Daniel L. **Population Genetics in Forensic DNA Typing**

Butler, J. (2005). **Forensic DNA Typing Biology, Technology, and Genetics of STR Markers** (2nd ed.). Burlington: Elsevier Science.

Thompson, Ewing, Frank, Pogemiller, Nolde, Koehler, . . . Storts. (2013). **Developmental validation of the PowerPlex® Y23 System: A single multiplex Y-STR analysis system for casework and database samples.** *Forensic Science International: Genetics, 7*(2), 240-250.

Forensic Biology Unit (2007). **SDPD AmplflSTR YFiler PCR Validation.**

Forensic Biology Unit (2012). **SDPD Forensic Biology Technical Manual.**

MFK. 2013. **Oregon State Police Forensic Services Division DNA Unit Y-STR Validation.**

Swgdam. Retrieved January 02, 2017, from http://www.swgdam.org/about-us

The FBI Quality Assurance Standards Audit For DNA Databasing Laboratories. (2011, September 01). Retrieved January 02, 2017 from http://media.wix.com/ugd/4344b0_c374a6d5edbd447e8eea5d7924ccbfb7.pdf

Yfiler Plus PCR Amplification Kit User Guide. (2014, October). Retrieved from https://tools.thermofisher.com/content/sfs/manuals/4485610_YfilerPlus_UG.pdf