2015

Determining the change in PCR efficiency with cycle number and characterizing the effect of serial dilutions on the DNA signal

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Thesis

DETERMINING THE CHANGE IN PCR EFFICIENCY WITH CYCLE NUMBER

AND CHARACTERIZING THE EFFECT OF SERIAL DILUTIONS ON THE DNA

SIGNAL

by

CHENG-TSUNG HU

B.S., National Chung Hsing University, 2008

Submitted in partial fulfillment of the

requirements for the degree of

Master of Science

2015

Approved by

First Reader _____

Catherine Grgicak, Ph.D
Assistant Professor of Biomedical Forensic Science

Second Reader _____

Robin Cotton, Ph.D
Associate Professor and Director of Biomedical Forensic
Science

ACKNOWLEDGEMENTS

DETERMINING THE CHANGE IN PCR EFFICIENCY WITH CYCLE NUMBER AND CHARACTERIZING THE EFFECT OF SERIAL DILUTIONS ON THE DNA SIGNAL

CHENG-TSUNG HU

ABSTRACT

The ability to obtain deoxyribonucleic acid (DNA) profiles is generally considered a powerful tool when examining evidence associated with a crime scene. However, variability in peak heights associated with short tandem repeats (STR) signal complicates DNA interpretation; particularly, low-template complex mixtures, which are regularly encountered during evidentiary analysis. In order to elucidate the sources that cause peak height variability a dynamic model, which simulates; 1) the serial dilution process; 2) polymerase chain reaction (PCR); and 3) capillary electrophoresis (CE) was built and used to generate simulated DNA evidentiary profiles.

In order to develop the dynamic model, PCR efficiencies were characterized. This was accomplished using empirical quantitative polymerase chain reaction (qPCR) data. Specifically, the ratios of fluorescent readings of two consecutive cycles were evaluated. It was observed that the efficiency fluctuated at early cycles; stabilized during the middle cycles; and plateaued during later cycles. The relationship between the change in efficiency and the concentration

of amplicons was modeled as an exponential function. Subsequently, this exponential relationship was incorporated into the dynamic model as a part of the PCR module.

Using the dynamic laboratory model, the effect of serially diluting a concentrated DNA extract to a low–template concentration was assessed in an effort to determine whether serially diluted samples are a good representation of evidence samples which contain low copy number of cells. To accomplish this, peak height variances and the frequency of drop-out between serially and non-serially diluted samples were compared. The results showed that diluting the sample had a substantial influence on allelic drop-out. However, the distributions of the observed peak heights did not consistently change; though, changes in peak height distributions became more pronounced with samples at lower targets. The peak height equivalency (PHE) was also used to aid in the determination of the effect of serial dilutions on reproducibility. There was not a major change in PHE between serially and non-serially diluted samples.

TABLE OF CONTENTS

LIST OF TABLES

targets. Analytical threshold of 50 RFU was applied.

PCR efficiency (σE) of 0% at locus D7S820 with final targets of 0.031 ng (Left Column) and 0.013 ng (Right Column). Two regression functions were used: Gaussian (Top), and Exponential (Bottom). ●: stock solution of 1000 ng/µL (subjected to large dilutions), ■: stock solution of 1 ng/µL (subjected to moderate dilutions), and ▲: variable (subjected to no dilution).

ABBREVIATIONS

AT   Analytical threshold

CE   Capillary electrophoresis

C_n The concentration of DNA amplicons at cycle n

C_0 The concentration of DNA molecules before amplification

ΔE   Change in PCR efficiency

dNTP   Deoxynucleotide triphosphates

DNA   Deoxyribonucleic acid

DO   Drop-out

f(DO)   Frequency of drop-out

$F_{E\cong 1}$ Fluorescent signal obtained when $E \cong 1$

F_n The fluorescent signal reading at cycle n

F_{n+1} The fluorescent signal reading at cycle n+1

μL   Microliter

ng   Nanogram

$C_{E=1}$ Nominal concentration

PHE   Peak height equivalency

E   PCR efficiency

PCR   Polymerase chain reaction

qPCR   Quantitative polymerase chain reaction

σE   Relative deviation of PCR efficiency

RFU   Relative fluorescent unit

STR   Short tandem repeats

1.0   Introduction

1.1   Challenges in Forensic DNA Analysis

The polymerase Chain Reaction (PCR) is widely used in many different disciplines, including biochemistry and molecular biology. It is considered a sensitive method because the target deoxyribonucleic acid (DNA) fragments exponentially increase during each PCR cycle. If binary forensic DNA comparisons are made, the peak height variability is of little consequence. The exception is when peak heights are in the stochastic range – the range where allelic drop-out is likely. If continuous interpretation methods are employed, then the peak height variability is of importance since the determination of the likelihood ratio depends upon the ability to characterize and use peak height information. In either case, validation of the interpretation method is required, and typically requires a set of mock evidence samples. Usually, these validation samples are extracted, quantified, mixed (if applicable), diluted and taken to be a good representation of low-template evidence samples. Here the assumption that these diluted samples result in the same 'stochastic' results and peak height variability is revisited.

In the forensic setting, short tandem repeats (STR) analysis is commonly employed. Typically, forensic biological samples undergo DNA extraction, quantification, amplification, and capillary electrophoresis (CE). After CE, profiles containing allelic information at selected loci are generated; and interpretation of the profile is conducted. Each stage of the testing has a certain level of variance

that could potentially affect the result of the STR analysis. It is not only the experimental variability that has to be considered, but also the nature of the evidentiary samples. Profiles obtained from evidence found at crime scenes are not pure and typically contain small quantities of DNA and signal interference from other contributors. This can lead to allele drop-out within a mixture profile. Since the condition of the evidentiary samples cannot be controlled, it is important to understand the effect of experimental variability on DNA signal and to determine if the experimental procedure needs to be modified in order to prevent the occurrence of lost information which leads to more difficult interpretation.

1.2 Dynamic Modeling

A dynamic model was previously developed to simulate the forensic DNA analysis process from quantification to capillary electrophoresis (CE)[1]. There were three modules in the model: serial dilutions (Figure 1.); PCR (Figure 2.); and CE (Figure 3.). In the serial dilutions module, up to three dilutions could be conducted, and the mean volume and standard deviation of the volume pipetted for DNA stock and Buffer could be modified. In addition, the DNA molecules were assumed to be evenly distributed, according to the Poisson distribution, within the solution[2]. In the PCR module, the model was built based on the loci used in Identifiler® PLUS STR amplification kit[1]. Here, the user has the ability to change the variation in PCR efficiency (E) where E is assumed to be normally distributed around mean 1. The CE module was developed in order to simulate the process

of DNA separation and allows the user to modify the relative standard deviation of peak height originating from CE. Therefore, a profile containing allelic information and allele height information is generated after every run or simulation.
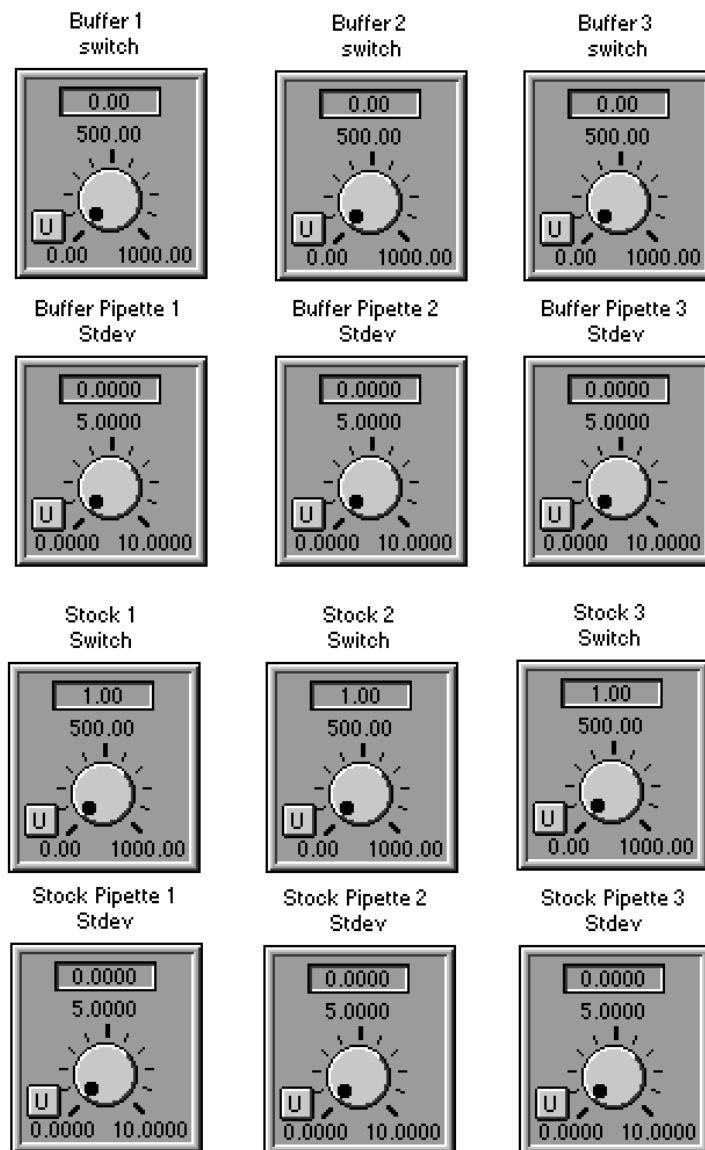
Figure 1. *The interface of the section of the serial dilutions of the dynamic model.*
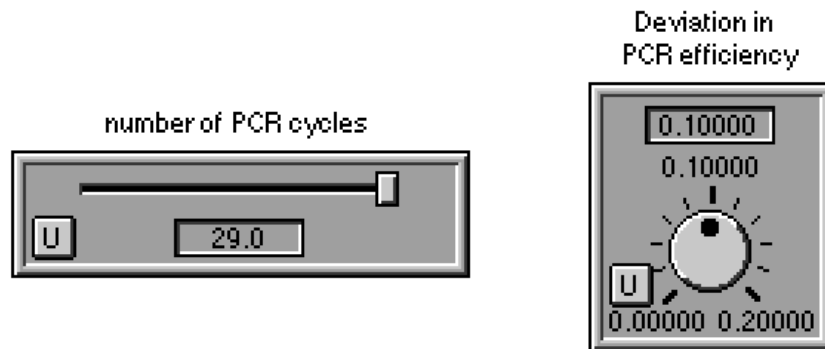
Figure 2. *The interface of the section of the polymerase chain reaction (PCR) of the dynamic model.*

	RSD of CE
RSD of CE[D8S1179 allele 1]	0.1
RSD of CE[D8S1179 allele 2]	0.1
RSD of CE[D21S11 allele 1]	0.1
RSD of CE[D21S11 allele 2]	0.1
RSD of CE[D7S820 allele 1]	0.1
RSD of CE[D7S820 allele 2]	0.1
RSD of CE[CSF1PO allele 1]	0.1
RSD of CE[CSF1PO allele 2]	0.1
RSD of CE[D3S1358 allele 1]	0.1
RSD of CE[D3S1358 allele 2]	0.1
RSD of CE[TH01 allele 1]	0.1
RSD of CE[TH01 allele 2]	0.1
RSD of CE[D13S317 allele 1]	0.1
RSD of CE[D13S317 allele 2]	0.1
RSD of CE[D16S539 allele 1]	0.1
RSD of CE[D16S539 allele 2]	0.1
RSD of CE[D2S1338 allele 1]	0.1
RSD of CE[D2S1338 allele 2]	0.1
RSD of CE[D19S433 allele 1]	0.1
RSD of CE[D19S433 allele 2]	0.1
RSD of CE[vWA allele 1]	0.1
RSD of CE[vWA allele 2]	0.1
RSD of CE[TPOX allele 1]	0.1
RSD of CE[TPOX allele 2]	0.1
RSD of CE[D18S51 allele 1]	0.1
RSD of CE[D18S51 allele 2]	0.1
RSD of CE[D5S818 allele 1]	0.1
RSD of CE[D5S818 allele 2]	0.1
RSD of CE[FGA allele 1]	0.1
RSD of CE[FGA allele 2]	0.1

Figure 3. *The interface of the section of the capillary electrophoresis (CE) of the dynamic model.*

The purpose of building this dynamic model was multi-fold: One of its main purposes is to allow the user to enter and modify various laboratory settings which may impact DNA signal. Once variability is introduced into the laboratory model, the simulated data could be used to evaluate the effects of these laboratory modifications on the signal. Thus, laboratory processes can be scrutinized individually with any experimental condition, without expending a lot of resources.

1.3   Variations of the Polymerase Chain Reaction (PCR) Efficiency

Many studies have examined, analyzed or modeled PCR efficiency while trying to properly determine its value. For example, there are two methods that have been used to evaluate the cycle efficiency; First is calculating the ratio between the number of DNA molecules between subsequent cycles [3,4]; Second is calculating the ratio between the fluorescence signal at a cycle and the fluorescence signals between subsequent cycles[5,6]. Higuchi et al. (1993), Gevertz et al. (2005) and Rutledge et al. (2008) suggested that the fluorescence signal at a cycle proportionally represent the number of DNA molecules at that cycle in a positive direction[5–7]. Therefore, the fluorescence signal obtained using quantitative polymerase chain reaction (qPCR) was chosen in this study to evaluate PCR efficiency because the fluorescence data for every cycle could be monitored and obtained through real-time qPCR.

As a result, though PCR efficiency is typically assumed to be 1 in;

$$C_n = C_0 \cdot (1 + E)^n \qquad \text{(Equation 1.)},$$

where C_n is the concentration of DNA amplicons at cycle n, C_0 is the concentration of DNA molecules before amplification; and E is the efficiency of amplification, stochastic variation associated with primer binding, polymerization and denaturation, makes this unlikely. Rather, it may be reasonable to assume that $E = 1 \pm \sigma E$, where σE represents variation in efficiency, which means that E is slightly different at every cycle. Further, PCR efficiency may not only vary around some mean E, but the mean E is known to decrease as cycle number increases. This phenomenon, observed as reaching a plateau, is well documented in both the forensic and biological literature[8,9]. The changing concentration of reagents and amplicons play an important role in this phenomenon[8,9]. Therefore, qPCR raw signal data was studied in order to explore the relationship between PCR amplicon numbers and the mean E. Once the relationship was established, the deviation of PCR efficiency and the change in PCR efficiency were incorporated into the model. To evaluate the effect of σE, four values, 0%, 5%, 10%, and 20%, were utilized. The use of the four values was included in the study since the real value of the variability is unknown.

1.4   The Effect of Serial Dilutions

Studies which examine the effects of serial dilutions can be seen in many types of studies including those which examined bacterial counts[10,11],

immunoassays[12] and STR signals[13]. In forensic settings, when samples with large quantities of DNA are tested, the profiles are likely to have off-scale allele peak heights and possible artifacts causing difficulty during interpretation. On the other hand, samples with small quantities result in the loss of information or, in the extreme case, sub-par comparisons. As a consequence, samples may need to be diluted to an optimal mass range so that the product of the amplification can result in a profile that leads to a correct interpretation. However, Liao and Duan suggested that it is necessary to take dilution error into consideration when calculating the original concentration of a sample after serial dilution[14]. Grgicak et al. (2010) also found that the serial dilutions can have a substantial influence on the calibration curves in a forensic setting[15]. The error generated by serially diluting a sample has also been studied within the context of statistics[16,17], microbiology[10,11], and epidemiology [12]. Despite the errors associated with the process, Kontanis and Reed (2006) suggested that the use of serial dilution could decrease the amount of inhibitors in the amplification, resulting in an increase in the PCR efficiency[18]. Given the fact that most validation studies consist of samples that have been serially diluted, it is of interest to evaluate its effects on peak height variability.

1.5   Pipetting Error and Sampling Error

Sampling error is a part of the dilution process as is pipetting error. Pipette error may originate from individual pipettes as well as individual operators. Pipette error has been incorporated into the dynamic model (Figure 1.) as

standard deviations and relative standard deviations, and were based on the ISO8655, values for acceptable accuracy and precision ranges of pipettes[19]. In addition, sampling error is also a source of peak variability. Sampling error is the variation that occurs when different aliquots of liquid contain varying numbers of copies of an allele. For instance, a 1 μL aliquot of liquid may contain 10 copies of allele 1 and 9 copies of allele 2. If the same extract was used for a second amplification, the second 1 μL aliquot may contain 8 and 12 copies of allele 1 and 2, respectively. These sampling variations have been studied before, and have been modeled with binomial and Poisson distributions[13,20,21]. If a sample has a low copy number of DNA, it is possible that the aliquot contains 0 copies of the allele. In this instance, allele drop-out is the result. However, Higgins et al. suggested that sampling error is relatively small compared to the dilution error and does not cause inaccuracy in the result of the immunoassay[12].

In the forensic field, since it is not-uncommon to encounter samples with low DNA quantity and quality, it is of interest to examine how serial dilutions affect the STR signal.

2.0   Characterizing Change in PCR efficiency (ΔE) with cycle number
2.1   Methods
2.1.1   Sample Preparation

Four previous run real-time quantitative polymerase chain reaction (qPCR) data sets using Quantifiler® Duo DNA Quantification kit (Applied Biosystems,

Foster City, CA) were obtained. The data included information for 60 samples

total. This included 16 samples from RD08-0006(062708CMG), 16 samples from

RD08-0005(012109CMG), 24 samples from QC013(030212CMG), and 4

samples from RD08-0006(103009CMG). The samples contained eight different

starting DNA amounts (50, 16.67, 5.56, 1.85, 0.617, 0.206, 0.069, 0.023 ng/µL).

All samples went through the qPCR thermal cycling process with denaturing

temperature of 95°C for 15 seconds, and annealing/elongation temperature of

60°C for 1 minute for 40 cycles. The fluorescent intensity at each cycle was

recorded, and the results were exported for further analysis.

2.1.2   Real-time quantitative polymerase chain reaction (qPCR)

The Quantifiler® Duo DNA Standard (200 ng/µL) was used to generate all

of the standard dilutions. Serial dilutions were conducted and 8 different starting

DNA concentrations (50, 16.67, 5.56, 1.85, 0.617, 0.206, 0.069, 0.023 ng/µL)

were obtained for the study. For each sample, the reagents of the reaction

included 12.5 µL of the PCR reaction mix and 10.5 µL of the primer mix along

with 2 µL of the DNA standard samples. The PCR reaction mix consisted of

AmpliTaq Gold® DNA Polymerase (Applied Biosystems, Foster City, CA) and

deoxynucleotide triphosphates (dNTP), and the primer mix consisted of primers,

probes and an passive reference control labeled with ROX^{TM} Dye. The

fluorescent dye does not emit light while it is coupled with the probe. During the

PCR step, the endonuclease activity of the Taq polymerase releases the

fluorescent dye from the probe, and as a consequence, the fluorescence is

detected. As the PCR reaction progresses, the fluorescence intensity increase since more of the fluorescent dyes are released from the probes. At the end of each cycle, a fluorescent signal reading was obtained using Applied Biosystems 7500 Real-Time PCR System with Sequence Detection Software v1.4.1; and it is this signal which was used to develop a relationship between fluorescence and amplicon concentration.

2.1.3 Relationship Between Fluorescence and Amplicon Concentration

The starting DNA concentrations were used to calculate the theoretical concentration at each cycle, and the theoretical concentration was determined by assuming the DNA concentration doubles during each round of qPCR. The polymerase chain reaction (PCR) efficiency at each cycle using fluorescent signals was also calculated using the following formula

$$Efficiency = \frac{F_{n+1}}{F_n} - 1 \qquad \text{(Equation 2.)},$$

where n was the cycle number, F_n was the fluorescent signal reading at cycle n, and F_{n+1} was the fluorescent signal reading at cycle $n+1$.

Another assumption is that the fluorescence proportionally represents the DNA amplicon quantity during the whole process[5–7]. Thus, when the value equals 1, the amount of DNA is doubling. Since a relationship between fluorescence and DNA amplicon concentration – which is unknown – was necessary, the *Efficiency* according to Equation 2. was calculated for all qPCR standards. Since the early phases of qPCR show no growth in fluorescence intensity – due to detection limits of the system – early points in the cycle were

discarded. Further, since plateauing is known to occur, the growth in fluorescence seen in later cycles were also discarded. Thus, to evaluate the relationship between measured fluorescence and amplicon concentration, only qPCR cycles with an efficiency closest to one were recorded. To estimate the "theoretical amplicon concentration", Equation 1. was utilized; with the assumption that the E is 1 in early phases of cycling. For example, if a sample with C_0 = 50 ng/µL resulted in an $E \cong 1$ at cycle 25, then that raw fluorescence at cycle 25 was recorded – plotted against the theoretical concentration of $50 \cdot 2^{25}$. In summary, by examining all available data in this way, a plot of fluorescence versus theoretical amplicon concentration was generated. An ordinary least squares regression ensued, and the slope was taken as the calibration sensitivity (i.e. the change in fluorescent signal with every 1 ng/µL increase in amplicon concentration).

2.1.4   Equation of The Change in Efficiency

After the relationship between fluorescence and amplicon concentration was established, the change in efficiency (ΔE) at each cycle was evaluated in an effort to establish the relationship between the concentration of the amplicon and the change in efficiency (ΔE). The change in efficiency (ΔE) was calculated at each cycle as

$$Change\ in\ Efficiency\ (\Delta E) = 1 - efficiency \qquad \text{(Equation 3.).}$$

When the value equals 0, it means the efficiency is 1, and there are no observable plateauing effects. It is the optimal value, and the value most

commonly assumed. When the ΔE value starts increasing, it suggests that the PCR efficiency is decreasing from 1, and the PCR is entering plateau phase. Therefore, to empirically determine the change in efficiency (ΔE) with number of amplicons, the ΔE was plotted against the theoretical number of amplicons for cycle n; where it was assumed that 1 diploid copy was 0.0063 ng of DNA. This resulted in 842 data points which were fit with an exponential function using Igor Pro software v6.36 (WaveMetrics, Inc.). The resultant equation was used in the PCR module of the dynamic model.

2.2   Results and Discussions

2.2.1   Fluorescence Versus Theoretical Amplicon Concentration

There are myriad studies that aim to characterize, model or discuss the PCR efficiency and the mechanisms by which it changes. One such study suggested that the cycle efficiencies were stable and close to 1 at early cycles of PCR, and starts decreasing toward 0 as cycle number increases[8]. Similarly, other work showed that the efficiency is constant; however, this was only possible when the PCR was using an optimized procedure[22]. If constant PCR efficiency is assumed, it might provoke inaccuracy in the PCR model. Thus, it is of interest to develop a PCR model free from these assumptions[23].

Therefore, in this study, the efficiencies at every cycle, for all samples, were calculated based on empirical data. The raw fluorescence data obtained from the qPCR were used. The results showed that the efficiencies were not

constant during the whole thermal cycling process and the values fluctuated at early cycles. This phenomenon has been seen in the literature using different calculation methods[6,24,25]. Therefore, the data at the early cycles was not used during analysis because the background noise was unpredicable[26].

The efficiencies then became stable and fell between 0 to 1 after cycle number 20 for samples with the largest initial amount concentration of DNA. The efficiency then started to decrease to 0 once sufficient cycle numbers were reached, and large numbers of amplicons were synthesized. Furthermore, it was observed that the cycle numbers where the efficiencies approached 1 increased as the initial amount of DNA decreased indicating that, as expected, the fluorescence was reaching detectable levels later in cycling when there were fewer initial copies of DNA. Not only is this result expected, but it has been observed previously[27]. For example, Bustin also observed this and suggested analyzing data with stable cycle efficiencies in order to gain more quantitative precision[28].

In this study, only the cycle with the efficiency closest to 1 was chosen from each sample, and the fluorescence value along with the projected amplicon concentration (calculated using Equation 1.) at that cycle from each data set as calculated by the methods described in Section 2.1.3 were included and graphed (Figure 4.).
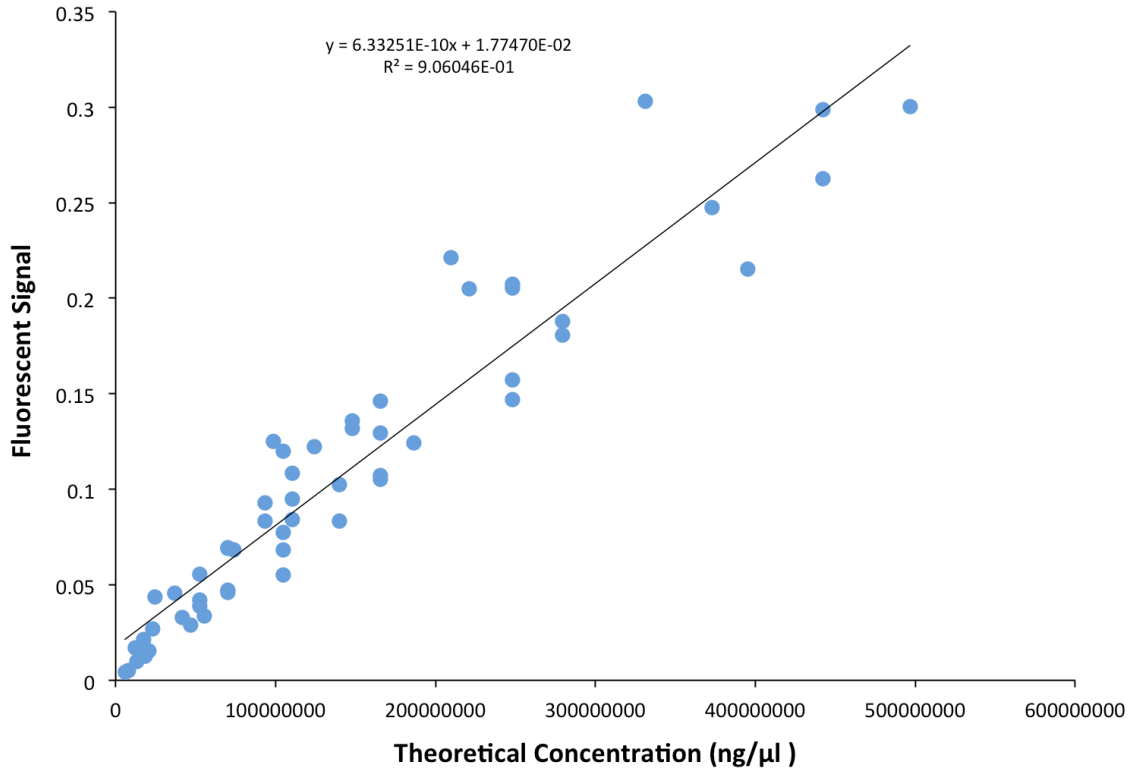
Figure 4. *Scatterplot of the theoretical concentration versus the fluorescent signal of qPCR data when efficiency ≅ 1.*

Qualitatively, it is observed that there is a strong linear correlation between the measured fluorescent signal and the expected concentration of amplicon at cycle$_{E=1}$. The graph was then fit with a line, and the following linear equation was generated:

$$F_{E \cong 1} = 6.33251 \times 10^{-10} C_{E=1} + 1.7747 \times 10^{-2} \qquad \text{(Equation 4.),}$$

where $F_{E \cong 1}$ was fluorescent signal obtained when E ≅ 1, and $C_{E=1}$ was nominal concentration (ng/µL) obtained via Equation 1.

This result was consistent with data published by Higuchi et al. who report that the fluorescence and amplicon concentration have a linear relationship[7].

This regression equation was used to calculate the actual DNA concentration at every cycle, and the data of the actual concentration was subsequently used for the change in efficiency (ΔE) study.

2.2.2 Change in Efficiency (ΔE) Versus Actual Amplicon Concentration

Many variations can affect PCR efficiency during the PCR process, including template concentration, depletion of reagents, and accumulation of the amplicons[29]. Other experimental designs may cause additional variation in the process as well. As a result, it has been suggested that the assumption of constant efficiency could lead to inaccurate results during PCR quantification[30]. Therefore, this study attempted to determine the change in efficiency (ΔE) during the end stages of thermal cycling in an attempt to estimate the drop in efficiency with amplicon concentration.

The change in efficiency (*ΔE*), calculated using Equation 3., for all cycles with appreciable fluorescence were plotted versus the nominal amplicon concentration which was calculated using Equation 4. The data was fitted with an exponential function (Figure 5.). The following regression equation was generated:

$$\Delta E = 1 - e^{-2.2533 \times 10^{-10} N} \qquad \text{(Equation 5.)},$$

where *ΔE* was the change in efficiency and *N* was the number of the amplicons generated during cycle n.
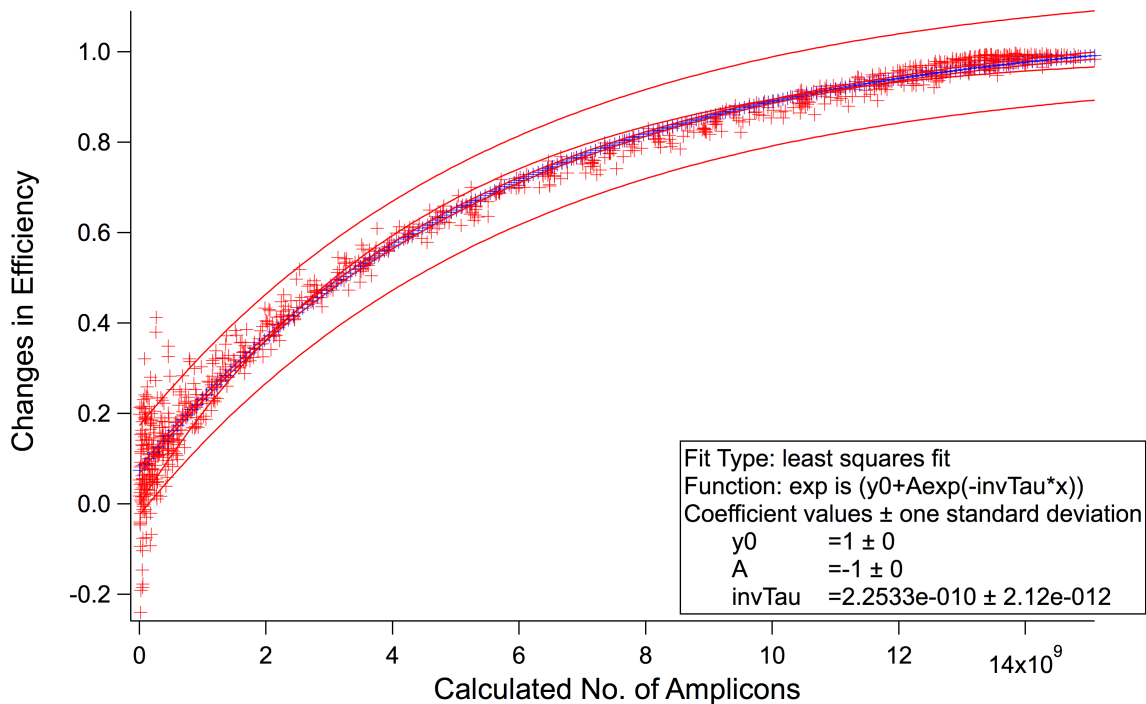
Figure 5. *Scatterplot of the change in PCR efficiency* (ΔE) *versus the nominal number of amplicons.*

The cycle efficiencies were found to be different at each cycle. Additionally, the change in efficiency has a larger variation with lower amplicon concentration. As the amplicon concentration increases, the change in efficiency also increases in an exponential fashion. In other words, the efficiencies decrease as more DNA molecules are generated. The change in efficiency approached 1 when the copy number of the DNA molecules was approximately $13*10^9$, indicating a significant number of amplicons are required before PCR plateau is reached.

This exponential relationship was subsequently implemented in the dynamic model of the forensic DNA process detailed in Section 1.2 of this thesis which was recently updated by Wellner, Rowan and Grgicak[31].

3.0   Characterizing the effect of serial dilutions on DNA signal

3.1   Methods

3.1.1   Simulated Data

In the dilution module of the dynamic model, three stock solutions were used: 1000 ng/µL, 1 ng/µL, and variable. The variable stock solutions were 0.1 ng/µL and 0.01 ng/µL. Profiles for the final target levels of 0.125 ng, 0.031 ng and 0.013 ng were generated. Therefore, the variable stock solutions did not require a serial dilution process, while data generated from the concentrated stocks of 1000 ng/µL and 1 ng/µL did. The 1000 ng/µL stock solution samples required larger dilutions and more dilution steps to obtain the final target than the 1 ng/µL stock solution sample. The dilutions used for each sample group are listed in Table 1.. In addition, pipette error was also incorporated into the model when pipetting was required. The standard deviations and relative standard deviations are listed in Table 2..

Table 1. *The serial dilutions used for different stock solutions and final target of the sample.*

Stock Solutions	Final Target = 0.125 ng	Final Target = 0.031 ng	Final Target = 0.013 ng
Variable	N/A	N/A	N/A
1 ng/µL	10:90	10:90→10:90	10:90→10:90
1000 ng/µL	1:99→1:99	1:99→1:99→10:90	1:99→1:99→10:90

Table 2. *The standard deviations and the relative standard deviations of the pipette volume used for the model.*

Pipetter Volume (µL)	Standard Deviation of the Pipette Volume	Volume into Amp (µL)	Relative Standard deviation of the Pipette Volume
1	0.14	1.25	11.2%
10	0.14	1.3	4.52%
90	0.85	3.1	10.8%
99	0.85		

In the PCR module, three possible amounts, which were 1.25 µL for the final target of 0.125 ng; 3.1 µL for the final target of 0.031 ng, and 1.3 µL for the final target of 0.013 ng, were transferred. In addition, the reaction was run for 32 dynamic cycles, which represents a 29-cycle amplification.

The relative deviation of PCR efficiency (σE) was varied. Values of 0%, 5%, 10%, and 20% were used. In addition, deviation during the capillary electrophoresis (CE) is also part of the dynamic model whereby the relative standard deviation of the CE set-up and injection was set to a constant 10%. The calibration sensitivities were locus dependent, and the values previously determined by Rowan[1] were utilized. Though the dynamic model can also

produce stutter peaks, noise signal and incorporate qPCR error, these were all set to 0 during this study.

The model was used to generate a hundred sample profiles for each final target for each stock solution for 0, 5, 10 and 20% of the σE; totaling 3600 dynamic runs. The output results included information on the known simulated genotype, the initial copy number input into the amplification for every allele value, and the corresponding peak height value in relative fluorescent unit (RFU). Thus, each profile contained 15 STR loci: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, and FGA, which represents a simulated AmpFLSTR® Identifiler® Plus profile obtained after a 29-cycle amplification run on a 3130 Genetic Analyzer[32].

3.1.2   The Frequency of Drop-out

Since one goal of the study was to assess the cause of allelic drop-out and the effect of the serial dilutions on frequency of drop-out, peak heights (in RFU) were compared the known simulated profile, and those alleles with a peak height of 0 RFU were identified. The alleles which exhibited a peak height of 0 RFU were then compared to the initial copy number, input into the amplification. The purpose of this was to elucidate the main cause of the peak height drop-out; and to determine whether allele drop-out was due to; 1) sampling; 2) PCR variations; or 3) a combination of both factors. All of those peak heights of 0 RFU were collected and sorted based on locus, relative deviation of PCR efficiency

(σE), target number, and stock solution concentration. The frequency of drop-out (f(DO)) was calculated as per the following equation;

$$f(DO) = \text{Number of alleles with 0 RFU}/ \text{Total number of alleles} \quad \text{(Equation 6.)}.$$

The peak height data was further filtered using analytical threshold of 50 RFU, and the number of alleles with peak height smaller than 50 RFU was obtained and used as the numerator for calculation of the f(DO), in order to assess the impact of the analytical threshold on f(DO).

3.1.3   Peak Height Analysis

All of the peak heights of 0 RFU described in the previous section were removed. The peaks with heights greater than 0 were sorted according to locus, relative deviation of PCR efficiency (σE), target mass, and the concentration of the stock solution. Histograms of peak heights were created using the auto-set bins function provided in Igor Pro (WaveMetrics, Inc.), and regression analysis was conducted using Gaussian and log-normal functions. The residuals were used to assess which regression functions resulted in the best fit to the data. After the distribution of the data was determined, the average peak height and standard deviation of the peak height were acquired from the coefficients of the regression. The goal was to assess the effect of serial dilutions on the average of peak heights, the spread, and to evaluate whether the distribution changed.

3.1.4   Peak Height Equivalency Analysis

The peak height data was also analyzed by calculating the peak height equivalency (PHE). To accomplish this, the peak height data, the same data set

that was examined in Section 3.1.3, was sorted based on locus, and the maximum peak height was identified. The peak height equivalency was calculated as per;

$$PHE = {peak\ height}/{maximum\ peak\ height} \qquad \text{(Equation 7.).}$$

Because of the way it is calculated, the PHE is always between 0 and 1. By comparing the PHE, the reproducibility can be assessed. If the experiments are reproducible, the PHE is expected to be close to 1. When the value of the PHE is close to 0, it means that reproducibility between amplifications is poor. Thus, if serially diluting highly concentrated stock solutions introduces additional variance, the PHE's spread would increase and/or the mean PHE would decrease. Therefore, normalized cumulative histograms were created using PHEs from every locus within the same relative deviation of PCR efficiency (σE) and target amount. The medians, and ranges for each group were determined and compared between data sets.

3.2   Results and Discussions

3.2.1   The Frequency of Drop-out

The peak heights of alleles of homozygous loci were removed from the data set such that only heterozygous loci were considered. In total, there were 85,684 heterozygous alleles examined. Of the 85,684 heterozygous alleles evaluated, 25,120 alleles dropped out. Of these, 25,119 alleles had zero copies at the initial cycle of PCR, indicating that the majority of the drop-out using the

model occurred due to sampling and only one drop-out occurred due to the insufficient amplification of an allele that was present (0.004%) (Table 3.). This corroborates the findings of Hedges who also suggested that the sampling error is the major source of the serial dilution variance[10]. However, it is contrary to Higgins et al. who stated that sampling error was shown to be negligible during immunoassay analysis[12]. Table 3. shows the frequency of drop-out obtained for each target level studied here, and like Higgins et al., shows that sampling effects predominately influence low-copy samples. Specifically, it is observed that almost all allelic drop-out events occurred when the target was less than or equal to 0.031 ng of DNA (i.e. ~5 copies of an allele).

Table 3. *Number of allele drop-out (DO) due to sampling and PCR at three targets. Data of each target included all relative deviations of PCR efficiency (σE), stock solutions and serial dilutions.*

	Target (ng)		
	0.125	0.031	0.013
Number of Allele DO Due to Sampling	26	12620	12473
Number of Allele DO Due to PCR	0	1	0
Number of Allele in Known Profile	28560	28644	28480

It should be noted that no analytical threshold was applied and thus, if a DNA molecule was present in the amplification, it was likely to result in a peak with a peak height > 1 RFU.

However, many forensic laboratories utilize an analytical threshold of 50 RFU. In this case, there were 27,603 alleles below threshold – an increase of 2,483 – suggesting that if an analytical threshold is applied, amplification variations impact the number of peaks which pass the detection threshold; despite the presence of the DNA molecule in the amplification tube. This suggests that when evaluating the source of allelic drop-out in a forensic profile, higher analytical thresholds result in larger drop-out rates, and that both sampling and PCR effects can influence these rates. However, the majority of drop-out is due to the total absence of a DNA template molecule in the amplification. Therefore, from a practical perspective, 64%, 0.8%, and 16% of alleles at 0.125 ng, 0.031 ng, and 0.013 ng sample target (Table 4.) could be detected with post-PCR enhancement such as larger injection times or post-PCR purification. In a similar vein, Kinnaman showed that post-PCR purification can increase the number of alleles detected by 76% and 100% for 5 second and 10 second injection time for the lowest DNA target of 0.0625 ng respectively[33]. While Smith and Ballantyne showed that the post-PCR purification cannot recover a full profile with DNA samples at 0.039 ng and below[34]. These studies corroborate this study's findings, which suggest that if an analytical threshold is used, and there is enough DNA present in the amplification, some signal can be regained by post-

PCR enhancements. However, the data presented here introduces the notion that post-PCR enhancement will not be able to recover the majority of signal from low-template samples which are subjected to sampling effects.

Table 4. *Number of allele drop-out at analytical threshold (AT) of 50 RFU and potential allele recovery (%)*

	Target (ng)		
	0.125	0.031	0.013
Number of Allele Drop-out at AT=50	72	12719	14812
The Potential Allele Recovery	64%	0.8%	16%

Frequency of drop-out (f(DO)) was also evaluated with data generated from utilization of diverse serial dilutions – the results are shown in Table 5. It was found that the effect of the serial dilutions was minimal when samples had a large amount of DNA (0.125 ng). However, the effect of serial dilutions became pronounced when smaller quantities of the DNA samples were assessed. The frequency of drop-out for samples without using serial dilutions was approximately 37%. The number increased to approximately 48% when serial dilutions were utilized, regardless of the relative deviation of the PCR efficiency (σE). Higgins et al. also concluded that the serial dilution error decreased when the number of target molecules increased[12]. Specifically, Higgins at el. evaluated serial dilutions error by simulating two immunoassay models that have dilution

error incorporated, and showed that the difference between the value of the true concentration and the simulated concentration is relatively small and consistent suggesting the higher the target, the more accurate the simulated data[12]. This is a substantial outcome as it suggests, DNA samples that were not subjected to serial dilutions always had a lower frequency of drop-out than those that had serial dilutions.

Interestingly, an increase in the relative deviations of PCR efficiency (σE) caused few differences in the frequency of drop-out (f(DO)), suggesting again that sampling effects play a major role in attaining full profiles.

Table 5. *Rate of drop-out (%) with different relative deviations of PCR efficiency (σE), stock solutions, dilutions, and targets. An analytical threshold of 0 RFU was applied.*

Target (ng)	Stock Solution (ng/µL)	Dilution	Relative Deviation of PCR efficiency (σE,%)			
			0	5	10	20
	1000	1:99→1:99	0.1	0.4	0.2	0.1
0.125	1	10:90	0.1	0.1	0.0	0.0
	0.1	N/A	0.0	0.0	0.0	0.0
	1000	1:99→1:99→10:90	48.4	51.2	47.8	46.9
0.031	1	10:90→10:90	44.9	46.3	47.9	46.9
	0.01	N/A	37.2	38.8	36.8	36.5
	1000	1:99→1:99→10:90	46.7	46.3	50.3	50.6
0.013	1	10:90→10:90	45.2	45.6	45.3	47.7
	0.01	N/A	36.1	38.6	37.0	36.4

The analytical threshold of 50 RFU was also applied to these data (Table 6.). The higher analytical threshold generally increased the rate of allele drop-out, but the trend among the data stayed the same as when no analytical threshold was applied, again suggesting that the application of an analytical threshold introduces additional loss of allelic information.

Table 6. *Rate of drop-out (%) with different relative deviations of PCR efficiency (σE), stock solutions, dilutions, and targets. An analytical threshold of 50 RFU was applied.*

Target (ng)	Stock Solution (ng/µL)	Dilution	Relative Deviation of PCR efficiency (σE,%)			
			0	5	10	20
0.125	1000	1:99→1:99	0.1	0.5	0.3	0.6
	1	10:90	0.1	0.2	0.1	0.6
	0.1	N/A	0.0	0.0	0.0	0.4
0.031	1000	1:99→1:99→10:90	48.4	51.2	47.8	48.2
	1	10:90→10:90	44.9	46.3	47.9	48.1
	0.01	N/A	37.2	38.8	36.8	38.0
0.013	1000	1:99→1:99→10:90	50.4	51.9	56.8	64.4
	1	10:90→10:90	49.1	50.4	53.7	61.3
	0.01	N/A	41.3	44.6	46.7	53.7

3.2.2   Peak Height

Since homozygous loci have overlapping allele peak height usually, a peak height of a homozygous locus is usually divided by two to represent the peak height of a single allele. Thus, it is assumed that each peak contributes equally to the total peak height observed in the electropherogram. Because of this limitation, it is difficult to find a fair method to determine how much of the homozygous peak should be attributed to which source. In order to avoid this limitation, all of the peak heights from homozygous loci were filtered and eliminated from the analysis. The remaining peak heights at all 15 loci were examined individually. In order to determine whether there is a single distribution that can describe the peak heights at all targets, the Gaussian and log-normal functions were fit to the data. A Gaussian function was applied to the data of final target of 0.125 ng, and the Gaussian curve fit the distribution of the data well (Figure 6. A). A log-normal function was also tested (Figure 6. B), and the outcome was similar.
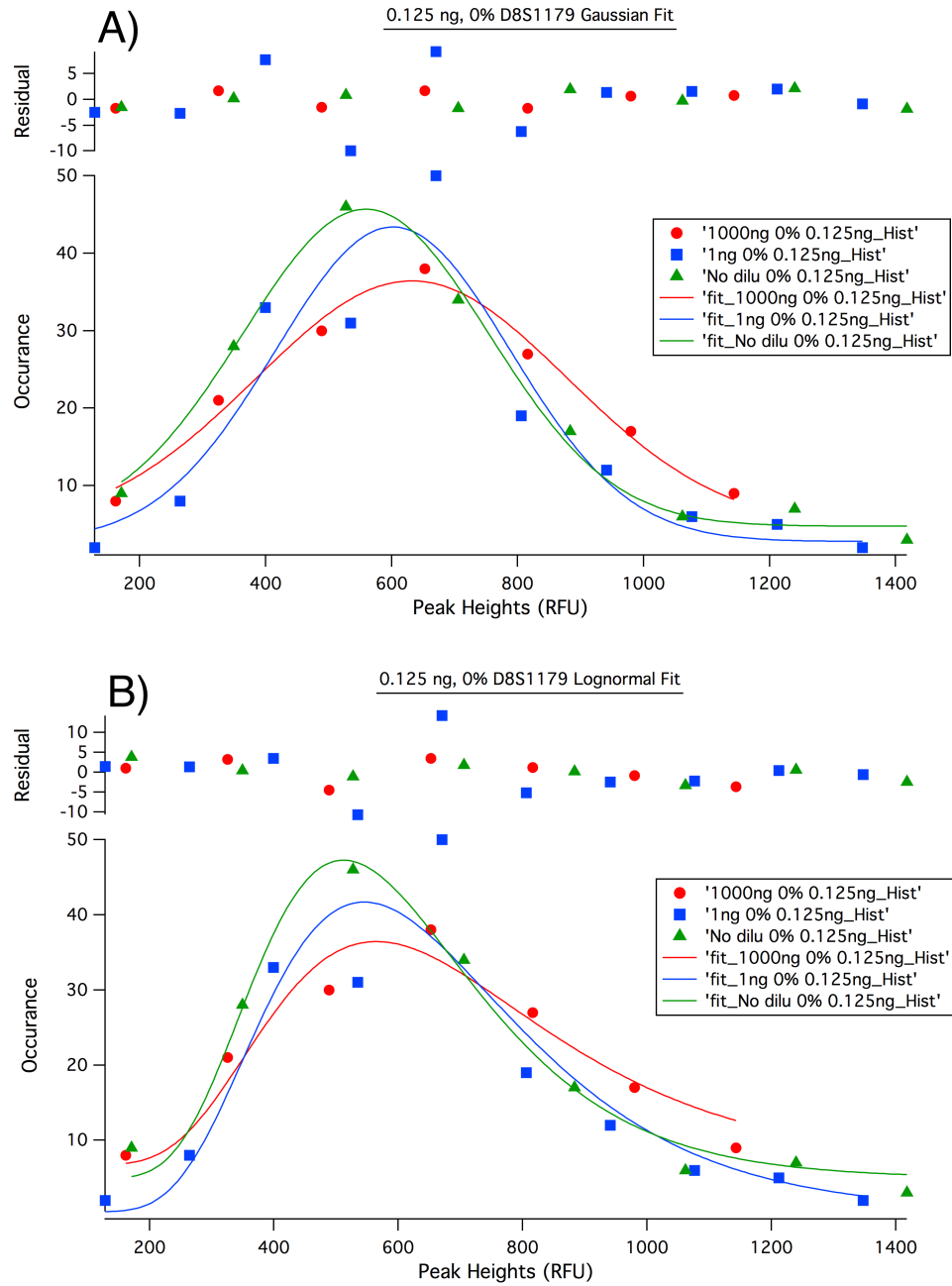
Figure 6. *Histograms of peak heights with relative deviation of PCR efficiency (σE) of 0% at 0.125 ng at locus D8S1179 with A) Gaussian Fits and B) Lognormal fits. Residuals of the fits are displayed on the top of each plot.* ●: *stock solution of 1000 ng/μL (subjected to large dilutions),* ■: *stock solution of 1 ng/μL (subjected to moderate dilutions), and* ▲: *variable (subjected to no dilution).*

In order to determine which function was a better fit for the data, the sum of the squared residuals were calculated. If a regression function is representing the data well, the residual will approach 0. Large values are indicative of poor fits.

The result of the sum of the squared residuals for 15 loci for Gaussian fits and log-normal fits at 0.125 ng with 0% σE are summarized in Table 7. Overall, the Gaussian fits had smaller values in the sum of the squared residuals than the log-normal fits. For instance, 12 out of 15 loci resulted in smaller residuals for Gaussian fits for the samples which underwent a large dilution. Similarly, 10 out of 15 loci for the medium and no dilution samples resulted in smaller residuals with Gaussian fits. These results indicate that the Gaussian function fits the peak height data better than the log-normal function, and the distribution of the peak heights was determined to be Gaussian. On the other hand, the result of the sum of the squared residuals at 0.125 ng with 20% σE (Table 8.) shows that the log-normal function became a better fit of the peak height data than the Gaussian function. The results for 0% and 20% σE show that different relative deviations of PCR efficiency (σE) could affect the class of distribution of peak height.

Table 7. *The sum of the squared residuals for 15 loci for the Gaussian and log-normal fits for samples at 0.125 ng with relative deviation of PCR efficiency of 0% (σE).*

Locus	Gaussian Fit			Log-normal Fit		
	Large Dilution	Medium Dilution	No Dilution	Large Dilution	Medium Dilution	No Dilution
D8S1179	14.805	301.619	17.228	59.201	369.855	36.288
CSF1PO	74.269	57.809	17.456	158.625	46.382	74.529
D7S820	385.834	134.099	37.021	192.945	336.777	160.375
D21S11	175.920	11.284	48.175	70.256	59.724	64.150
D3S1358	6.433	188.778	167.920	55.548	79.092	82.944
TH01	113.445	63.076	96.873	120.512	68.059	45.155
D13S317	116.786	63.483	88.388	94.838	114.760	42.448
D16S539	43.846	109.226	32.026	161.640	48.674	103.437
D2S1338	11.821	88.684	122.105	95.275	86.608	117.032
D19S433	61.850	31.784	87.498	308.964	88.520	149.831
vWA	23.898	38.666	273.501	91.007	158.947	283.650
TPOX	81.870	8.458	70.616	101.150	11.124	86.526
D18S51	51.828	91.352	65.527	228.706	167.312	79.744
D5S818	11.621	56.302	40.532	88.201	43.952	20.319
FGA	50.342	47.799	80.087	171.042	131.847	106.622

Table 8. *The sum of the squared residuals for 15 loci for the Gaussian and log-normal fits for samples at 0.125 ng with relative deviation of PCR efficiency of 20% (σE).*

Locus	Gaussian Fit			Log-normal Fit		
	Large Dilution	Medium Dilution	No Dilution	Large Dilution	Medium Dilution	No Dilution
D8S1179	17.716	178.955	88.772	11.829	52.941	7.016
CSF1PO	85.015	83.720	83.368	16.932	11.252	75.590
D7S820	156.712	152.064	6.535	98.067	79.648	44.526
D21S11	219.538	47.051	108.242	22.555	181.010	70.410
D3S1358	54.674	38.533	84.023	10.876	9.622	42.586
TH01	84.195	78.422	50.784	30.541	24.341	9.987
D13S317	32.909	81.754	68.283	15.679	33.378	16.099
D16S539	106.771	116.287	62.171	694.602	93.289	28.028
D2S1338	24.383	375.784	334.402	74.038	78.235	104.863
D19S433	96.129	55.643	93.698	49.073	23.096	22.832
vWA	92.605	97.078	56.136	44.631	18.279	38.003
TPOX	25.349	30.622	13.195	20.350	18.008	32.825
D18S51	128.738	11.704	14.568	47.373	103.917	78.956
D5S818	71.987	30.934	133.698	3.1079	27.176	77.567
FGA	104.331	134.051	41.182	30.824	24.214	13.355

However, when peak height distribution analysis was attempted with the data obtained from low-template samples, (e.g., 0.031 ng and 0.013 ng samples), neither the Gaussian, nor log-normal functions could successfully be fit to the data. Examination of D7S820 shown in Figure 7. shows that no apex in peak heights is reached. This result is observed for most of the data. That is, for low-template samples, the number of peak heights in the lower ranges continuously increase, resembling an exponential growth rather than a Gaussian

or log-normal distribution. The only low-template samples that resulted in an apex were those originating from samples not subjected to dilutions.
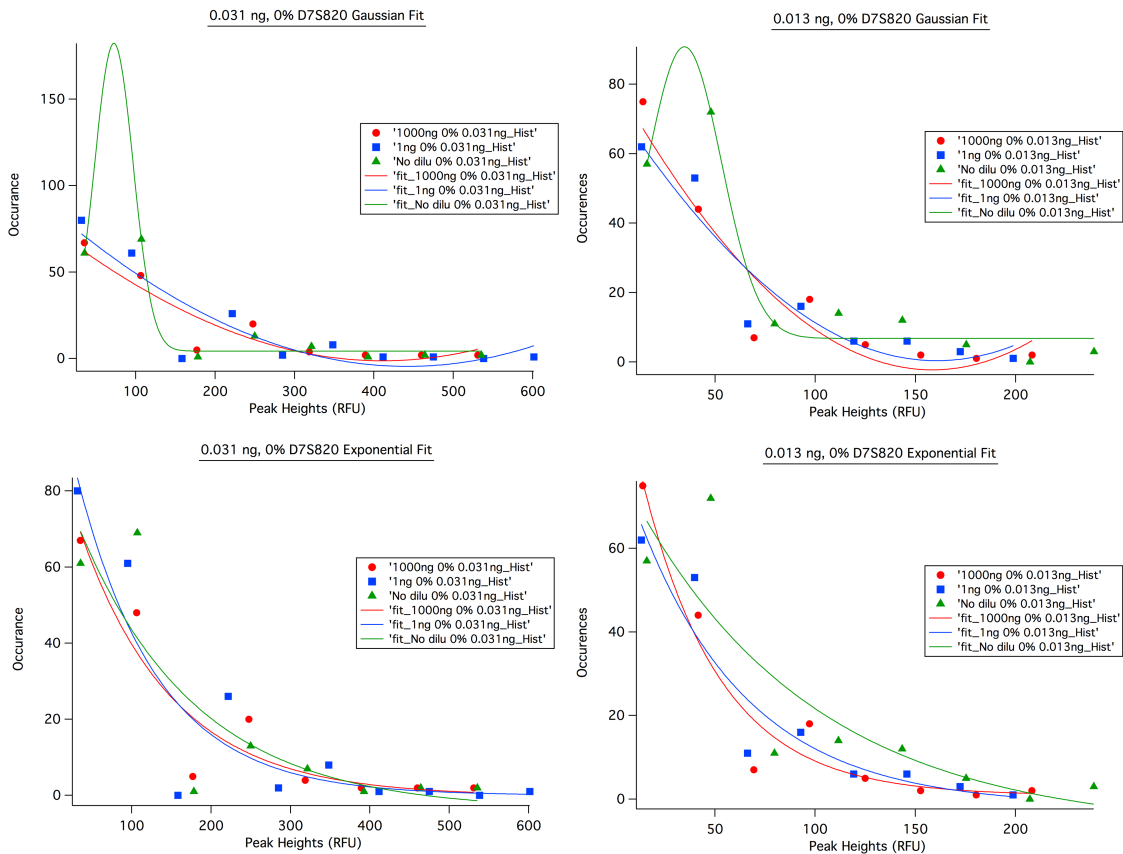
Figure 7. *Histograms of peak heights with relative deviation of PCR efficiency (σE) of 0% at locus D7S820 with final targets of 0.031 ng (Left Column) and 0.013 ng (Right Column). Two regression functions were used: Gaussian (Top), and Exponential (Bottom). ●: stock solution of 1000 ng/µL (subjected to large dilutions), ■: stock solution of 1 ng/µL (subjected to moderate dilutions), and ▲: variable (subjected to no dilution).*

As a result, only the 0.125 ng samples were used to assess; 1) whether the Gaussian versus log-normal are good representations of peak height distributions, and 2) whether dilutions impact the distributions. The distributions of the peak heights at 0.125 ng for 15 loci were evaluated for each locus individually. One representative locus from each color channel was chosen and the data is shown in Figure 8.

Furthermore, Table 7. and 8. show the values of the sum of the squared residuals for each locus for samples that underwent significant, moderate and no dilution. However, there were no consistent changes in these values between different levels of dilution at all 15 loci, indicating the distribution of the peak heights does not change with dilution-level.
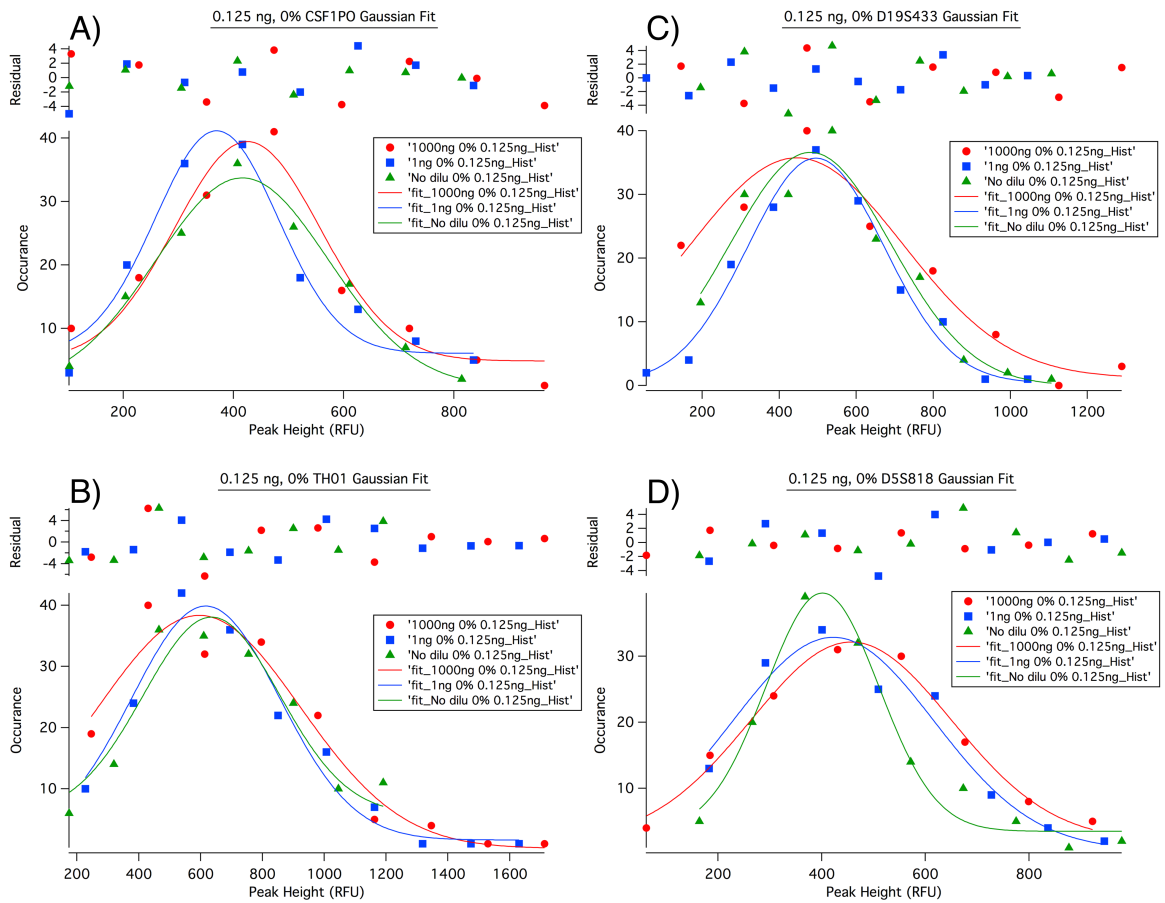
Figure 8. *Histograms of peak heights with relative deviation of PCR efficiency of 0% (σE) at 0.125 ng at Locus A) CSF1PO, B) TH01, C) D19S433 and D) D5S818. Residuals of the fits are displayed on the top of each plot. ●: stock solution of 1000 ng/µL (subjected to large dilutions), ■: stock solution of 1 ng/µL (subjected to moderate dilutions), and ▲: variable (subjected to no dilution).*

The averages and standard deviations were collected from the coefficient values of the Gaussian fits. The averages of the peak heights at all 15 loci were graphed using a category plot, where the error bars represent +1 and −1 standard deviation (Figure 9.). The graph shows that serially diluting the samples

did not have any visible, nor consistent, effect on the peak height averages. Further, there seemed to be no visible, nor consistent, impact on the standard deviation of the peak height. For instance, samples that did not undergo dilution had smaller standard deviation at 8 out of 15 STR loci, when compared to samples that were diluted 1,000-fold. In conclusion, subjecting samples to large dilutions did not result in changes in the properties of the peak height signal.
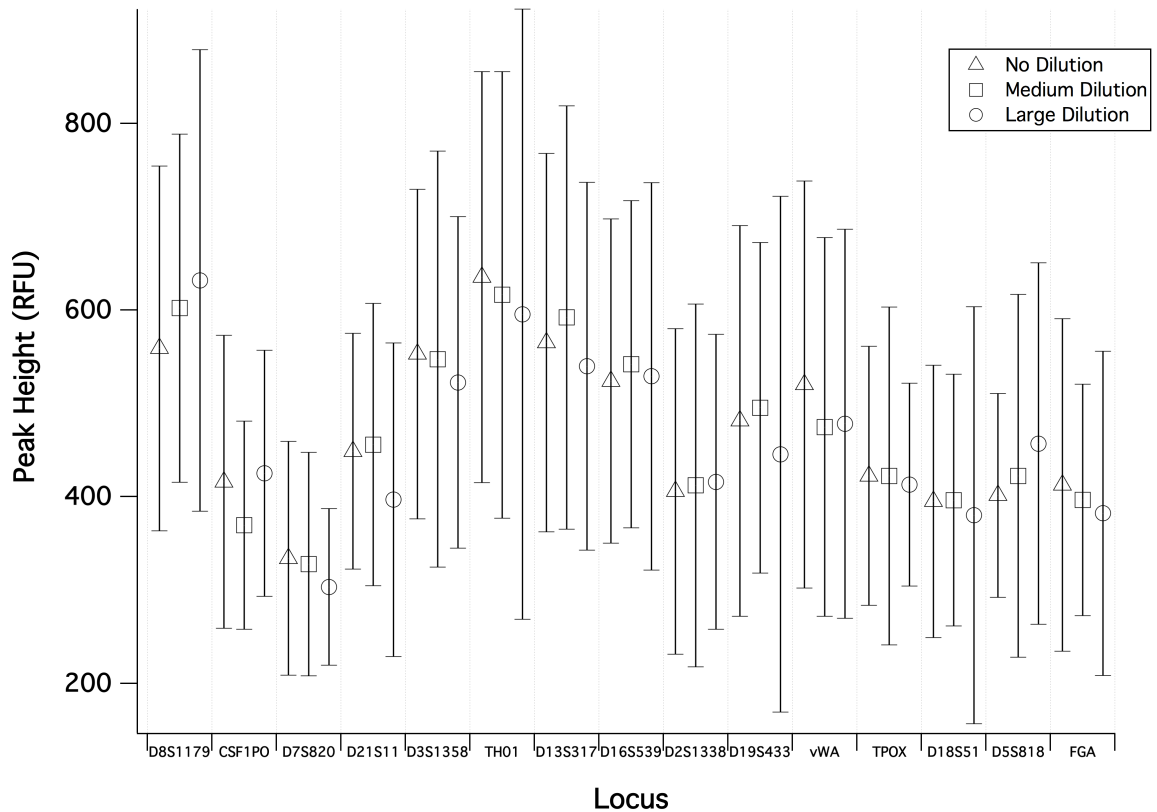
Figure 9. *Category plot of the average of peak height of the 15 loci with relative deviation of PCR efficiency (σE) of 0% at the final target of 0.125 ng. Error bars showed the +1 and -1 standard deviation.*

3.2.3   Peak Height Equivalency

Peak height Equivalency (PHE) was determined as described by Equation 7. in Section 3.1.4 for every locus for the 36 data sets. There were a total of 100 samples, per data set. The data were generated according to the inputs provided in Table 9. and 10.. The data sets included 4 different relative standard deviations of PCR efficiency (σE, 0%, 5%, 10% and 20%), 3 Targets (0.125 ng, 0.031 ng and 0.013 ng) and 3 different dilution processes (the large dilution, the medium dilution and the no dilution).

Once determined, the PHEs from the different loci within the same data set were determined. The means, standard deviations, medians, minimums and maximums were identified for every data set. All data are summarized in Table 9. and Table 10.

Table 9. *Calculated mean, standard deviation, median, minimum, maximum and range of the PHE with 0% and 5% relative deviation of PCR efficiency (σE).*

σE (%)	Target (ng)	Stock Solution (ng/µL)	Mean	Standard Deviation	Median	Min	Max	Range
0	0.125	1000	0.412	0.185	0.391	0.040	0.999	0.959
		1	0.443	0.176	0.429	0.043	0.997	0.954
		0.1	0.460	0.175	0.445	0.052	0.994	0.942
	0.031	1000	0.329	0.169	0.244	0.116	0.970	0.854
		1	0.330	0.170	0.246	0.141	0.997	0.856
		0.01	0.329	0.179	0.239	0.125	0.993	0.868
	0.013	1000	0.301	0.160	0.238	0.093	0.996	0.903
		1	0.327	0.176	0.247	0.093	0.991	0.898
		0.01	0.315	0.166	0.234	0.099	0.995	0.896
5	0.125	1000	0.399	0.177	0.379	0.035	0.994	0.959
		1	0.418	0.174	0.396	0.039	0.999	0.960
		0.1	0.414	0.170	0.397	0.055	0.971	0.916
	0.031	1000	0.317	0.173	0.255	0.100	0.950	0.850
		1	0.337	0.169	0.270	0.134	0.996	0.862
		0.01	0.350	0.184	0.271	0.126	0.983	0.857
	0.013	1000	0.332	0.177	0.271	0.092	0.997	0.905
		1	0.323	0.175	0.262	0.101	0.995	0.894
		0.01	0.311	0.172	0.249	0.078	0.995	0.917

Table 10. *Calculated mean, standard deviation, median, minimum, maximum and range of the PHE with 10% and 20% relative deviation of PCR efficiency (σE).*

σE (%)	Target (ng)	Stock Solution (ng/µL)	Mean	Standard Deviation	Median	Min	Max	Range
10	0.125	1000	0.320	0.163	0.289	0.021	0.967	0.946
		1	0.342	0.161	0.312	0.020	0.997	0.977
		0.1	0.371	0.170	0.347	0.036	0.987	0.951
	0.031	1000	0.295	0.177	0.237	0.053	0.991	0.938
		1	0.321	0.191	0.258	0.054	0.977	0.923
		0.01	0.319	0.184	0.260	0.086	0.994	0.908
	0.013	1000	0.294	0.177	0.238	0.060	0.991	0.931
		1	0.301	0.176	0.247	0.041	0.997	0.956
		0.01	0.296	0.175	0.249	0.059	0.995	0.936
20	0.125	1000	0.217	0.154	0.177	0.009	0.996	0.987
		1	0.241	0.162	0.199	0.011	0.995	0.984
		0.1	0.254	0.169	0.211	0.016	0.981	0.965
	0.031	1000	0.210	0.162	0.160	0.018	0.970	0.952
		1	0.216	0.160	0.166	0.024	0.981	0.957
		0.01	0.229	0.174	0.177	0.022	0.997	0.975
	0.013	1000	0.227	0.162	0.181	0.026	0.975	0.949
		1	0.256	0.199	0.192	0.018	0.997	0.979
		0.01	0.200	0.155	0.159	0.015	0.983	0.968

Overall, the PHE did not change when different levels of dilution were applied to the sample at the same target. For example, when the medians are evaluated, there is no consistent trend observed. That is, the PHE did not consistently increase, or decrease with dilution-level across the four σE data sets. Looking at the ranges of PHE's (Table 9. & 10.), samples with the same σE and same target had no characteristic trend that could be found when different

dilution processes were utilized. For example, a trend, the samples with the largest dilution had the smallest range and samples without serial dilution had the largest range, was found with samples with 0% σE at 0.031 ng; however, the opposite trend was also found within samples with 0% σE at 0.125 ng and 0.013 ng, samples with 10% σE at 0.031 ng, and samples with 20% σE at 0.125 ng and 0.031 ng.

Further, when the PHE's of the data sets for different σEs are compared, the ranges of PHEs for samples with 0% σE and 5% σE are indistinguishable; however, the ranges of PHE increase as the σE increases to 10% and 20%. Furthermore, σE's affected the PHE median, where the medians decrease as the σE increases; an expected result given the PHE is a measure of peak height reproducibility and the peak height is expected to be affected by the PCR process.. Lastly, though the PHE's tend to decrease as the DNA template mass decreasess, the differences are less striking when σE is large (i.e. 10% and 20%).

4.0   Conclusions

The aim of this study was to characterize two variations in the forensic deoxyribonucleic acid (DNA) analysis: one was the change in cycle efficiencies of the PCR, and the other was the serial dilution effect on DNA signal. For the cycle efficiency, there are two sources of variation; one come from random error; and another results from the change in efficiency, which decreased with an increase

in amplicon production. Real time quantitative polymerase chain reaction (qPCR) raw fluorescence signal was used to characterize the change in efficiency, which was later incorporated into the dynamic model. For the second part of the study, the effect of serial dilutions was examined.

To evaluate the change in efficiency, the qPCR efficiencies were determined by calculating the ratio of the florescence signals of two consecutive cycles. It was observed that the polymerase chain reaction (PCR) fluorescent signal was not stable at the early stages of thermal cycling. The signal became stable after a sufficient quantity of amplicon was produced. The number of cycles required for to reach signal stability depended on the starting amount of DNA, where the signal passed the detection threshold later in cycling if there was smaller levels of starting material. Once the signal was deemed measurable, PCR efficiencies were determined by examining the increase in signal between two successive cycles. In early cycles, the efficiency was $\cong$ 1. In later cycles, PCR efficiencies started to decrease because the amount of the DNA molecules became large and plateauing effects became detectable. Comparison of the raw fluorescence signal and the theoretical concentration of DNA showed that the florescence signal had a direct linear relationship with the nominal copy number of the DNA molecules. This proportionality was in turn used to establish the actual copy number of the DNA molecules after each thermal cycle based on the observed fluorescence. A plot of the change in efficiency versus amplicon number showed that PCR efficiency decreased in an exponential manner when

compared to number of amplicons. This result was incorporated into the dynamic model.

In the study of characterizing the effect of serial dilutions, the source of the drop-out, the frequency of drop-out (f(DO)), peak height, and peak height equivalency were evaluated. First, it was found that the majority of the peaks consisting of 0 RFU resulted from sampling error, and only one drop-out originated from insufficient amplification. Secondly, the result of the f(DO) showed that the effect of the serial dilution was not potent with samples at higher target. However, the effect of serially diluting the sample became recognizable with samples amplified using lower DNA targets. However, when an analytical threshold was applied, 64%, 0.8%, and 16% of the alleles that dropped out at 0.125 ng, 0.031 ng, and 0.013 ng were associated with amplification effects. This suggests that though the majority of allele loss is due to sampling effects, a small, but substantial level of alleles may be recovered by using smaller analytical thresholds (or no analytical thresholds) or enhanced post-PCR processes. Thirdly, an evaluation of the peak height showed that the distribution of the peak height for samples amplified using 0.125 ng with 0% σE was well represented with a Gaussian class of distribution. This was determined based on the calculation of the sum of the squared residuals. Interestingly, the distribution of the peak heights for samples with 0% σE at 0.031 ng and 0.013 ng could not be described by a Gaussian or log-normal distribution and may be better represented by an exponential relationship. For the 0.125 ng samples, the

distribution class did not change between samples that were serially diluted and those that were not. However, the distribution changed to log-normal class with samples at 0.125 ng with 20% σE, suggesting that the PCR variation can affect the class of peak height distribution. In addition, the averages and standard deviations of the peak height at 0.125 ng did not change between samples that were serially diluted and those that were not. Lastly, analysis of the peak height equivalency (PHE) showed that serially diluting concentrated DNA samples did not consistently affect peak height reproducibility. However, when PHE was evaluated, the median PHE increased with a decrease in the relative deviation of PCR efficiency (σE); and was larger for the 0.125 ng sample than with the other two low-template targets tested, suggesting both PCR variation and sampling effects can impact peak height reproducibility.

In conclusion, it was determined that the serial dilution does have an effect on the frequency of drop-out; and is mainly caused by sampling error. Further, as long as the template DNA is present during the amplification process, the effect of serial dilutions does not substantially affect the peak height distribution class. Further, when PHE was examined, serially diluting a sample did not consistently impact the PHE medians, suggesting the reproducibility of peak heights is artificially decreased by utilization of dilutions. These results suggest that in order to obtain representative drop-out rates from validation samples, the use of serial or large dilutions is not recommended post-extraction, or post-quantification. Rather, the forensic validation may be best served by generating data from

samples where the whole liquid (i.e. whole blood, whole saliva) has been diluted

prior to cell lysis.

LIST OF JOURNAL ABBREVIATIONS

Anal Biochem	Analytical Biochemistry
Biophys J	Biophysical Journal
Biotechnol Bioeng	Biotechnology and Bioengineering
BMC Biotechnol	BMC Biotechnology
BMC Mol Biol	BMC Molecular Biology
Chem Eng Sci	Chemical engineering science
Forensic Sci Int Genet	Forensic Science International: Genetics
Int J Food Microbiol	International Journal of Food Microbiology
J Chemom	Journal of Chemometrics
J Forensic Sci	Journal of Forensic Sciences
J Mol Endocrinol	Journal of Molecular Endocrinology
Nat Biotech	Nature Biotechnology
Neurosci Lett	Neuroscience Letters
Nucleic Acids Res	Nucleic Acids Research
Proc Natl Acad Sci	Proceedings of the National Academy of Sciences

References

1. Rowan KE. Characterizing variability in fluorescence-based forensic DNA measurement and developing an electrochemical-based quantification system [Internet]. 2014;Available from: http://search.proquest.com/docview/1560245408?accountid=9676

2. Stenman J, Orpana A. Accuracy in amplification. Nat Biotech 2001;19(11):1011–2.

3. Peccoud J, Jacob C. Theoretical uncertainty of measurements using quantitative polymerase chain reaction. Biophys J 1996;71(1):101–8.

4. Stolovitzky G, Cecchi G. Efficiency of DNA replication in the polymerase chain reaction. Proc Natl Acad Sci 1996;93(23):12947–52.

5. Rutledge R, Stewart D. A kinetic-based sigmoidal model for the polymerase chain reaction and its application to high-capacity absolute quantitative real-time PCR. BMC Biotechnol 2008;8(1):47.

6. Gevertz JL, Dunn SM, Roth CM. Mathematical model of real-time PCR kinetics. Biotechnol Bioeng 2005;92(3):346–55.

7. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reactions. Nat Biotech 1993;11(9):1026–30.

8. Nordgård O, Kvaløy JT, Farmen RK, Heikkilä R. Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: The balance between accuracy and precision. Anal Biochem 2006;356(2):182–93.

9. Butler JM. Advanced Topics in Forensic DNA Typing: Methodology. Waltham, MA: Academic Press, 2012;

10. Hedges AJ. Estimating the precision of serial dilutions and viable bacterial counts. Int J Food Microbiol 2002;76(3):207–14.

11. Hedges A. Estimating the precision of serial dilutions and colony counts: contribution of laboratory re-calibration of pipettes. Int J Food Microbiol 2003;87(1–2):181–5.

12. Higgins KM, Davidian M, Chew G, Burge H. The Effect of Serial Dilution Error on Calibration Inference in Immunoassay. Biometrics 1998;54(1):19–32.

13. Timken MD, Klein SB, Buoncristiani MR. Stochastic sampling effects in STR typing: Implications for analysis and interpretation. Forensic Sci Int Genet 2014;11(0):195–204.

14. Liao JJZ, Duan F. Calibrating the concentration from a serial dilution process. J Chemom 2006;20(6-7):294–301.

15. Grgicak CM, Urban ZM, Cotton RW. Investigation of Reproducibility and Error Associated with qPCR Methods using Quantifiler® Duo DNA Quantification Kit*. J Forensic Sci 2010;55(5):1331–9.

16. Chase GR, Hoel DG. Serial Dilutions: Error Effects and Optimal Designs. Biometrika 1975;62(2):329–34.

17. Gelman A, Chew GL, Shnaidman M. Bayesian Analysis of Serial Dilution Assays. Biometrics 2004;60(2):407–17.

18. Kontanis EJ, Reed FA. Evaluation of Real-Time PCR Amplification Efficiencies to Detect PCR Inhibitors. J Forensic Sci 2006;51(4):795–804.

19. ThermoElectron Corporation. Finnpipette® focus single channel variable & fixed volume instructions for use. 2008;

20. Gill P, Curran J, Elliot K. A graphical simulation model of the entire DNA process associated with the analysis of short tandem repeat loci. Nucleic Acids Res 2005;33(2):632–43.

21. Perlin MW, Sinelnikov A. An Information Gap in DNA Evidence Interpretation. PLoS ONE 2009;4(12):e8327.

22. Ramakers C, Ruijter JM, Deprez RHL, Moorman AF. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. Neurosci Lett 2003;339(1):62–6.

23. Alvarez M, Vila-Ortiz G, Salibe M, Podhajcer O, Pitossi F. Model based analysis of real-time PCR data from DNA binding dye protocols. BMC Bioinformatics 2007;8(1):85.

24. Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N. Statistical significance of quantitative PCR. BMC Bioinformatics 2007;8(1):131.

25. Rutledge RG. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. Nucleic Acids Res 2004;32(22):e178–e178.

26. Ruijter JM, Ramakers C, Hoogaars WMH, Karlen Y, Bakker O, van den Hoff MJB, et al. Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. Nucleic Acids Res 2009;37(6):e45–e45.

27. Platts AE, Johnson GD, Linnemann AK, Krawetz SA. Real-time PCR quantification using a variable reaction efficiency model. Anal Biochem 2008;380(2):315–22.

28. Bustin S. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 2000;25(2):169–93.

29. Booth CS, Pienaar E, Termaat JR, Whitney SE, Louw TM, Viljoen HJ. Efficiency of the Polymerase Chain Reaction. Chem Eng Sci 2010;65(17):4996–5006.

30. Rutledge R, Stewart D. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. BMC Mol Biol 2008;9(1):96.

31. Wellner G, Rowan K, Hu C-T, Grgicak CM. Developing a Dynamic Model of the DNA Laboratory Process to Characterize the Sources of Uncertainty in DNA Signal: Applications to Forensic DNA Education, Training, and Validation. In: the American Academy of Forensic Sciences 67th Annual Scientific Meeting. Orlando, FL: American Academy of Forensic Sciences, 2015;

32. Applied Biosystems. AmpFlSTR Identifiler Plus PCR Amplification Kit User Guide. 2014;

33. Kinnaman EA. Evluating the efficacy of post-PCR purification columns using low template single source DNA amplified with Identifiler and Identifiler PLUS. 2013;

34. Smith PJ, Ballantyne J. Simplified Low-Copy-Number DNA Analysis by Post-PCR Purification. J Forensic Sci 2007;52(4):820–9.

Curriculum Vitae

Cheng-Tsung Hu
Year of Birth: 1986
617.901.6838
cthu@bu.edu

Education

- National Chung Hsing University, Taichung, Taiwan
  B.S. Soil and Environmental Sciences, 09/2004 – 06/2008

Laboratory Experience

- 05/2013 –10/2013, 06/2014 – Present Thesis Research, Thesis Adviser: Prof. Catherine M. Grgicak, Boston U., MA, Title: Determining The Change in PCR Efficiency With Cycle Number and Characterizing The Effect of Serial Dilutions on DNA Signal
- 06/2012 – 07/2012 Summer Internship, Supervisor: Prof. Catherine M. Grgicak, Boston U., MA, Topic: Finding an optimum analytical threshold for different ratios of mixtures for forensic DNA analysis
- 09/2007 – 06/2008 Member of Special Topic Research, Supervisor: Prof. Tan, Chen-Chung, Laboratory of Soil Ecology, NCHU Poster Presentation Topic: Enumeration of coliforms in soils applied with different fertilizers
- 06/2007 – 09/2007 Prof. Chang, Chia-Che's Research Team, Laboratory of Cell Physiology, NCHU, Taiwan
- 06/2006 – 06/2007 Participant of "Long-Term Ecosystem-Scale Project in the Wuling Area：Sources of Nitrate" (Shei-Pa National Park Headquarters, the Construction and Planning Agency, the Ministry of the Interior, 96). Supervisor: Prof, Peng, Tsung-Ren, Laboratory of Hydrogeology, NCHU

Presentation

- 02/2013 Christine A. Rakay, Joli Bregu, Cheng-Tsung Hu and Catherine M. Grgicak, Maximizing Allele Detection by Selecting Optimal Analytical Thresholds, American Academy of Forensic Science, Washington DC.
- 02/2015 Genevieve Wellner, Kayleigh Rowan, Cheng-Tsung Hu and Catherine M. Grgicak, Developing a Dynamic Model of the DNA Laboratory Process to Characterize the Sources of Uncertainty in DNA Signal: Applications to Forensic DNA Education, Training and Validation, American Academy of Forensic Science, Orlando, FL.