

EFFECTS OF DNA COPY NUMBER ON SHORT TANDEM REPEAT STUTTER
RATIOS

By

SAADIA AMIR KHAN

A thesis submitted to the

Graduate School-Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of Master of Science

Graduate Program in Chemistry

Written under the direction of

Dr. Catherine M. Grgicak

And approved by

Dr. Catherine M. Grgicak, Ph.D., M.S.F.S

Dr. David Salas-de la Cruz, Ph.D., EIT

Dr. George Kumi, Ph.D.

Camden, New Jersey

October 2021

THESIS ABSTRACT

Effects of DNA copy number on short tandem repeat stutter ratios

by SAADIA AMIR KHAN

Thesis Director:

Dr. Catherine M. Grgicak

Determining the weight of evidence against a suspect is best served by high fidelity signal representations of forensically relevant genomic regions of interest: Short tandem repeats (STR). During PCR, STR is a synthesized fragment one repeat unit shorter than wildtype. When electrophoresed and detected, the relative peak heights at the stutter and allele positions can be used as a proxy for the relative level of stutter to allele product generated during PCR, informing forensic DNA interpretation.

Since many forensic samples contain only a few copies of DNA, this study explores if relative stutter abundances are of the same distributions when originating from high- and low- copy numbers. Data analysis began by categorizing the signals as noise, stutter, and allele. Only stutter and allelic peaks were retained for further analysis. Samples were amplified at 0.25 ng and 0.0313 ng; data from 5 STR loci were interrogated, resulting in analysis across 620 data points. The data was first explored through visual representation via boxplot.

Before proceeding to statistical analysis, a comprehensive literature review demonstrated that the stutter models solely relying on per-locus descriptions of stutter ratios (SRs) may be improved by considering the sequence of the STRs themselves. As such, all statistical evaluations were conducted on a per-allele basis. Next, a t-test (p-value threshold: 0.005) was employed to explore if average SRs between high- and low-copy numbers are similar. The results show significantly different mean values between the two templates. This has far-reaching forensic implications since it suggests that classical binary approaches to genotype inference for low-template samples ought not utilize boundaries developed from high-template ones.

Subsequently, we tested the SRs between low- and high- template samples by evaluating the similarity of the distributions themselves. To complete this test, a Kolmogorov-Smirnov test was performed. Mostly, the distributions were distinct, wherein low template samples demonstrated a higher proportion of stutter peaks exceeding expectation. These results are similar to the t-test results, therein suggesting a more refined approach to genotype inference is required for complex forensic signal containing both high- and low- template levels within a single mixture.

Acknowledgments

I would like to express my sincere gratitude to my advisor, Dr. Catherine M. Grgicak, for providing me an opportunity to conduct research with her. Under her invaluable guidance, I was able to develop skills and learn methodologies to accomplish my thesis. Her dedication, dynamism, and motivation were an inspiration throughout my research work. It was a great privilege and honor to work and study under her guidance.

I would like to thank my thesis committee members, Dr. David Salas-de la Cruz and Dr. George Kumi, for their encouragement, insightful comments, and suggestions throughout my research and especially, the write-up process.

I would like to thank everyone at Rutgers for standing by me through this process. Special thanks to Nidhi and my fellow researchers.

I am extremely thankful to my lovely mom and aunt for their prayers, sacrifices, and encouragement. My heartfelt thanks to my family and friends: Ali, Adnan, Zia, Umara, Anoosha and Lubana. I have no words to say but thanks to my wonderful husband Faisal Khan for his unconditional support, understanding, and care. Special thanks and love to my little genius Aali Faisal for his understanding and sacrifice in sharing his mom with Rutgers.

Table of Contents

1. Introduction.....	1
1.1. Short Tandem Repeats (STR) and Forensic DNA Analysis	1
1.2. PCR Amplification, Allele Detection and Electropherogram.....	3
1.3. Stutter	8
2. Material and Methods.....	12
2.1. Data Preparation	12
2.2. Stutter	13
2.2.1. Stutter ratio distributions	13
2.2.2. Stutter Ratio analysis for mean values.....	14
3. Results and Discussion	15
3.1. Allele specific stutter ratios versus locus specific stutter ratios.....	15
3.1.1. TH01	20
3.1.2. D18S51	21
3.1.3. FGA.....	21
3.1.4. D2S1338	22
3.2. Stutter Ratio between high and low template samples	22
3.2.1. Reverse Stutter.....	24
3.2.1.1. FGA (20).....	24
3.2.1.2. FGA (23).....	27
3.2.2. Forward Stutters:	32

4. Conclusion	33
Appendix 1	34
TH01 (6)	34
TH01 (9)	35
D2S1338 (17).....	36
D2S1338 (21).....	37
Bibliography.....	38

List of Figures

Figure 1. An illustration of STR repeat units on the base of length of repeat units. The repeat units are classified as di-, tri-, tetra-, penta-, and hexa-nucleotide units, based on the number of bases within a repeat unit.....	1
Figure 2. An example of simple, compound, and complex STR tetranucleotide motifs.	
(b) An example of the STRs of two alleles at the same genetic location of two chromosomes. The STR genotype of this individual is G = 6, 7.....	2
Figure 3. A representative electropherogram (EPG) obtained from a single source 0.25 ng sample amplified with Identifiler™ Plus Amplification Kit, a forensically relevant assay consisting of 16 tetranucleotide STR loci. The EPG consists of four dye channels where peak heights in RFU are on the y-axis and the base-pair size of the DNA fragments are plotted along the x-axis. (X) indicates stutter peaks.....	6
Figure 4. A representative electropherogram presenting high and low template same source samples for locus D18S51. The blue arrows and numbers represent reverse stutters, while green arrows and numbers represent forward stutters. A ₁ and A ₂ (red color) represent allele 1 and allele 2 of locus D18S51.....	7
Figure 5. Schematic illustration of slipped strand model. During the replication process the DNA polymerase attaches itself to 3' end of the DNA and synthesizes identical complementary strands. Sometimes the strands dissociate, and the DNA polymerase complex becomes unpaired, and the replication stops. When the two strands pair up again, sometimes they mis-pair to create a shorter copy or, more rarely, a larger copy. These faulty copies are called stutters, and produce stutter or shadow peaks that are, typically, one STR unit smaller or larger than the biological allele.	9

Figure 6. Box plots of high template stutter ratios against both alleles of five loci FGA (red), D18S51(yellow), D2S1338(green), D13S317(dark green), TH01(light green). Both alleles of each locus are expressed by same color box plots. Box plots show the median, interquartile range, and outer quartiles (black dots).....	17
Figure 7. Scatter plot between difference in stutter ratio of both allele and difference in base pair length of a STR locus.....	19
Figure 8. Reverse Stutter: (a) A histogram (blue) of allele 20(FGA) with base pair size 225 bp for the high template samples. (b) A histogram (red) of allele 20(FGA) for the low template samples. (c) A KS- test plot showing cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.835$ and $p = <0.0001$. (d) A box plot 20(FGA) of all non-zero stutter ratios which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value 4.56 and p-value 0.0009.....	24
Figure 9. Reverse Stutter: (a) A histogram (blue) of allele 23(FGA) with base pair size 237 bp for the high template samples. (b) A histogram (red) of allele 23(FGA) for the low template samples. (c) A KS-test plot of cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.620$ and $p = 0.0114$. (d) A box plot 23(FGA) of all non-zero stutter ratios, which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value of 3.8 and p-value 0.0013.....	27

List of Tables

Table 1. A summary of p-value obtained from KS-test and t-test, the ranges and the proportions of zero-stutter values for five loci tested across the high template and low template category.....	29
Table 2. A summary of SRT obtained from average stutter ratios, standard deviation values for low and high template samples per- allele across five loci.....	31
Table 3. A summary of p-value obtained from KS-test and t-test, the ranges, and the proportions of zero-stutter values for 2 loci tested across high and low template category.	32

1. Introduction

1.1. Short Tandem Repeats (STR) and Forensic DNA Analysis

Short tandem repeats (STR) are the genetic markers most frequently used in forensic DNA analysis. These polymorphic DNA sequences are scattered throughout the human genome; however, their distribution among the chromosomes is not uniform. STRs are found within non-coding part of the human genome and they constitute approximately 3% of the entire genome¹. As illustrated in Figure 1, forensically relevant short tandem repeats – also known as microsatellites – consist of repeat units that are generally 2 - 6 base pairs in length, with the number of repeats varying among individuals. It is the variation among the human population that makes them effective genetic markers for human identification². On the base of length of the repeat units they are classified as mono-, di-, tri-, tetra-, penta- etc. In forensic DNA analysis, STRs with 3 - 5 nucleotide repeat units are often used.

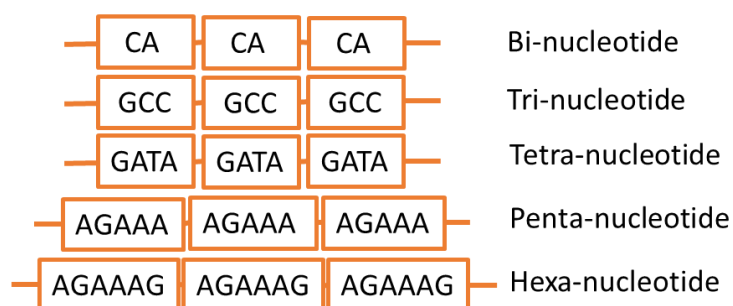


Figure 1. An illustration of STR repeat units on the basis of length of repeat units. The repeat units are classified as di-, tri-, tetra-, penta-, and hexa-nucleotide units, based on the number of bases within a repeat unit.

The STR repeat unit type also varies on the basis of repeat patterns, which is divided into three major categories as shown in figure 2 (a): i) simple repeats, consisting of an STR of identical length and sequence, for example, D5S818's $[AGAT]_n$; ii) compound repeats consisting of two or more simple repeat units, for example, vWA's $[TCTA] [TCTG]_n [TCTA]_n$; and iii) complex repeats consisting of multiple repeat blocks of variable unit lengths that interrupt the sequence with variable intervening sequences³, for example, D21S11's $[TCTA]_n [TCTG]_n [TCTA]_n TA [TCTA]_n TCA [TCTA]_n TCCA TA [TCTA]_n$.⁴

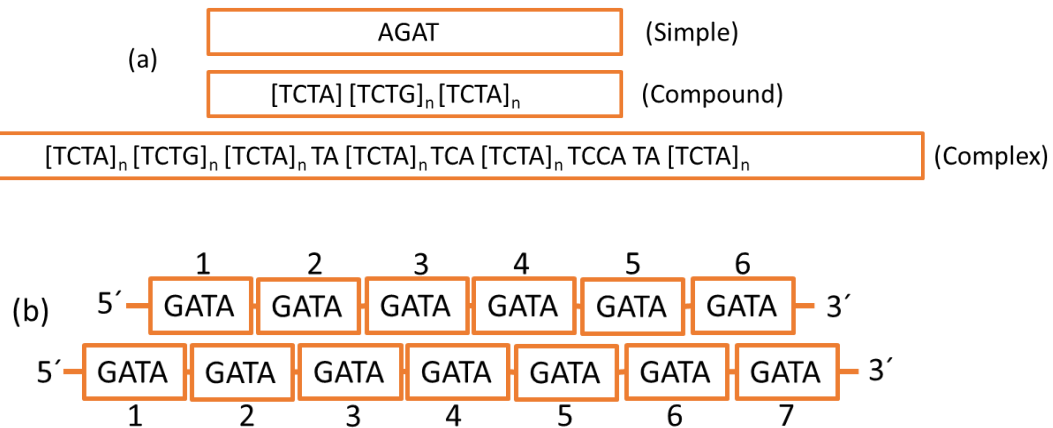


Figure 2. An example of simple, compound, and complex STR tetranucleotide motifs. (b) An example of the STRs of two alleles at the same genetic location of two chromosomes. The STR genotype of this individual is G = 6,7 (based on numbers of STR units).

Regardless of motif or base pair length of the STR, their hypervariability within the human population means they can be used to distinguish between individuals. Since each pair of chromosomes is a combination of two genes; one from the parental and one from the maternal sources, STR lengths can be used to compare the genotype of the biological source left at the crime scene and the genotype of a suspect. For example, in

Figure 2 (b) is an example of a person whose STR genotype is $G = 6,7$. Thus, if the crime scene's biological fluid also renders $G = 6,7$ then the forensic expert would report that the suspect could not be excluded as the source of biological material left at the scene.

1.2. PCR Amplification, Allele Detection and Electropherogram

In order to fully realize the potential made possible by this person-to-person genetic diversity contained within STR regions, one must be able to detect and size the DNA. The Polymerase Chain Reaction (PCR) is the common means by which to do this. PCR amplification is a technique used to synthesize hundreds of millions of copies of targeted DNA sequences within hours⁵. It involves the heating and cooling of DNA samples (typically, 0.5 ng - 1 ng) over a pre-defined cycle number, where one cycle includes the following three temperature steps: denaturation, annealing, and extension. Denaturation usually takes place at 95°C and causes complementary DNA strands to separate from each other. During the annealing step, the temperature is lowered to ca. 60°C which allows the primers (short oligonucleotide strands of ca. 25 nucleotides in length) to bind to their complementary DNA target sequences. In the extension step, the temperature is raised to 70-72°C allowing the DNA polymerase to catalyze the formation of new phosphodiester bonds.⁶ A PCR process under ideal conditions (100% efficiency) can produce $t_o \cdot 2^n$ copies of DNA over n PCR cycles where t_o is the initial number of DNA copies. The PCR process, however, is not perfect as it generates artifacts which complicates data interpretation.⁷ One notable PCR artifact is that of stutter which is the topic of this work.

Once amplification is completed, capillary electrophoresis ensues, and is used to detect the amplicons generated during the PCR process. Since DNA fragments are negatively charged, a potential difference across two electrodes is used to inject the DNA

into a capillary. Once injected, the potential continues to be applied such that the negatively charged PCR fragments travel from the cathode, through a polymer filled capillary, to the anode. The fragments are therein separated by length because shorter fragments travel more quickly through the polymer. A laser and a detector are present at the end of the capillary, and when the fluorophores tagged to the PCR primers are bombarded by the laser light, fluorescence results which is recorded for analysis. Since each primer is tagged with a known fluorophore, and the primer sequences are designed to hybridize to a specific locus (i.e., DNA location) of interest, the color indicates what location on the genome is being interrogated. In addition, since the fluorophores are tethered to the primers, and the primers are necessarily part of all newly synthesized fragments, the fluorescent intensity is a proxy of the number of fragments synthesized during PCR, which is, in turn, a representation of t_o .⁸ Thus, in an electropherogram (the end product of electrophoresis) the size of the fragment is represented on the x-axis and is deduced using the length of time it took the fragment to migrate through the capillary, and the peak height is a proxy of the number of synthesized fragments produced or targeted. Figure 3 is a representative image of such an electropherogram. In this representation, the electropherogram consists of four dye channels blue, green, yellow, and red originating from FAM, VIC, NED, and PET, fluorophores.⁹ Each channel consists of multiple loci, and each locus consists of multiple peaks. At each locus, large peaks representing a person's alleles are obvious when the template mass, t_o , is large. Notably, the genotype within a locus may be heterozygote, i.e., consisting of two alleles, or a homozygous where the STR alleles inherited from mother and father are identical. In addition to the intense peaks, multiple small peaks are also observed at each locus and represent the artifact commonly encountered in STR analysis

known as stutter. In samples like the one in Figure 3, stutter is comparably small to the allele peak and can, typically, be easily recognized as non-biological. Previous work has demonstrated that when t_o is low the stutter intensity can reach levels that make it difficult to distinguish it from allele. Thus, in this work, we examine the impacts of template mass on the relative stutter peak height.

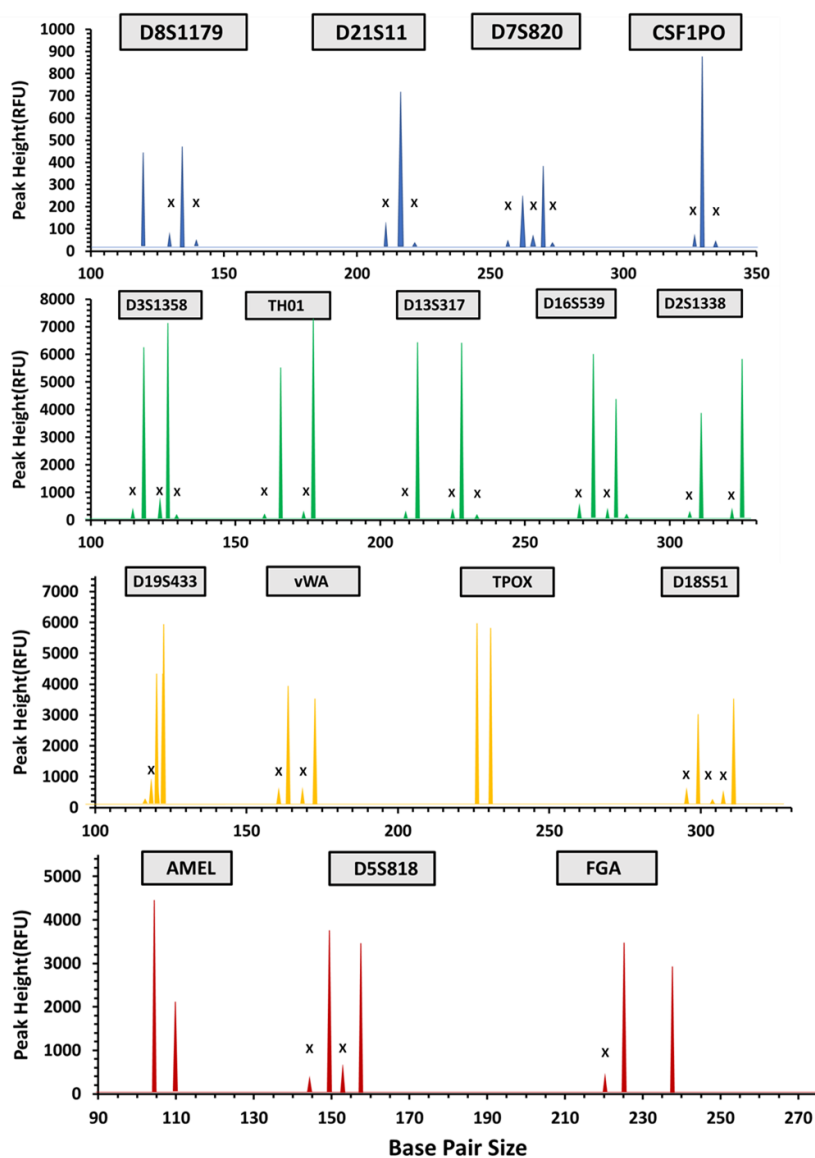


Figure 3. A representative electropherogram (EPG) obtained from a single source 0.25 ng sample amplified with Identifiler™ Plus Amplification Kit, a forensically relevant assay consisting of 16 tetranucleotide STR loci. The EPG consists of four dye channels where peak heights in RFU are on the y-axis and the base-pair size of the DNA fragments are plotted along the x-axis. (X) indicates stutter peaks.

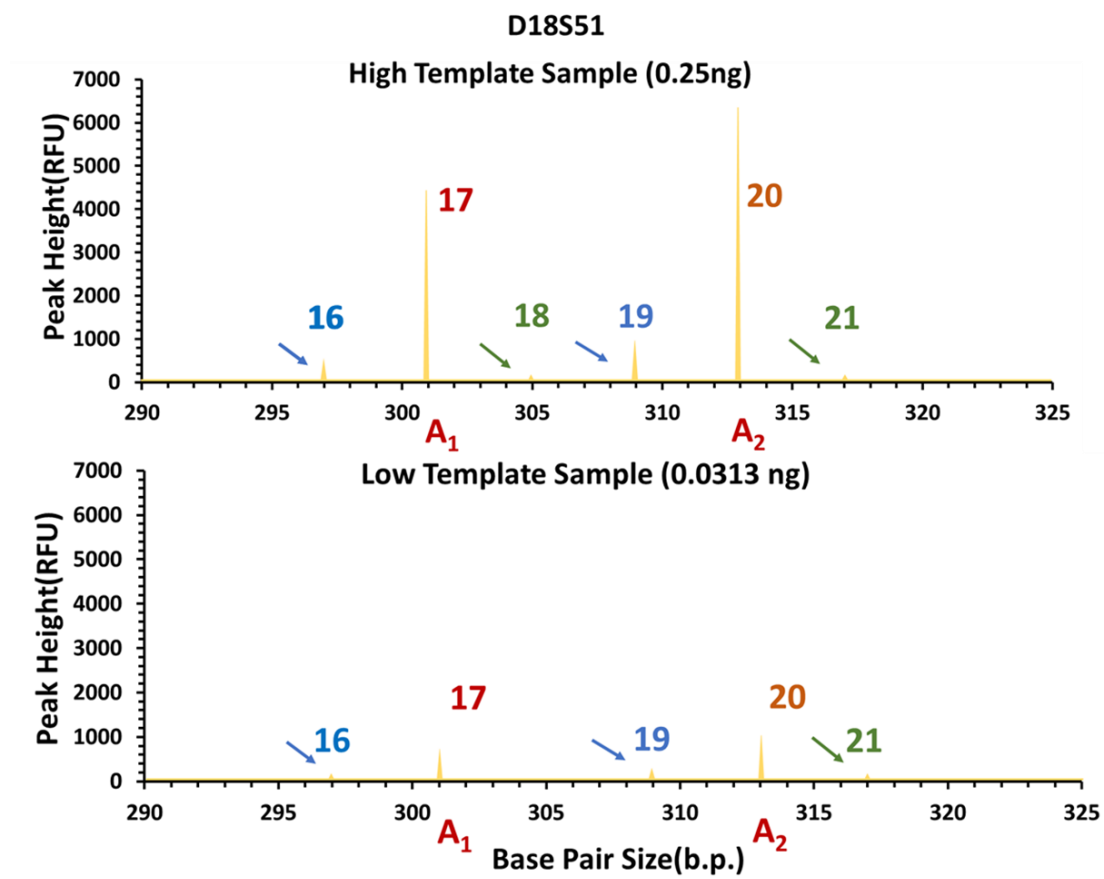


Figure 4. A representative electropherogram presenting high and low template same source samples for locus D18S51. The blue arrows and numbers represent reverse stutters, while green arrows and numbers represent forward stutters. A₁ and A₂ (red color) represent allele 1 and allele 2 of locus D18S51.

1.3. Stutter

Stutter is a PCR artifact that is often seen as minor peak one repeat unit smaller than the target STR allele product. This is known as “backward”, “back” or “reverse” stutter. Less frequently, a stutter product that is one repeat larger is produced. These minor peaks are often referred as “forward” or “plus” stutter.^{3, 10} Stutter products have been reported in the literature since STRs were first studied, and the primary mechanism by which to explain the in vitro synthesis of stutter products is referred to as the slipped-strand mispairing model.¹¹ As per the slipped-strand mispairing model, during the replication process a region of primer-template complex becomes denatured, which allows the formation of a loop, as depicted in Figure 5. If a repeat unit bulges out of the newly synthesized strand, then a single STR unit is inserted in the next amplification step to produce forward stutter product. If the repeat unit loops out in the template strand, it results in PCR products that are one repeat unit smaller than the template strand.¹²

Since stutter can occur at any cycle in the amplification, stutter intensities necessary vary; that is, strands that slip early in the PCR will produce more stutter since the newly formed stutter product will be copied in the same way as the full-size allele strand will. So, in cases where there are only a few templates strands, if one stutters early in cycling, a great proportion of the product is expected to be of the stuttered variety. Put concretely, if the total template copy number is 2, and in the first cycle one of the strands stutters then we obtain one stutter strand and three full sized strands at the end of this cycle. In the next

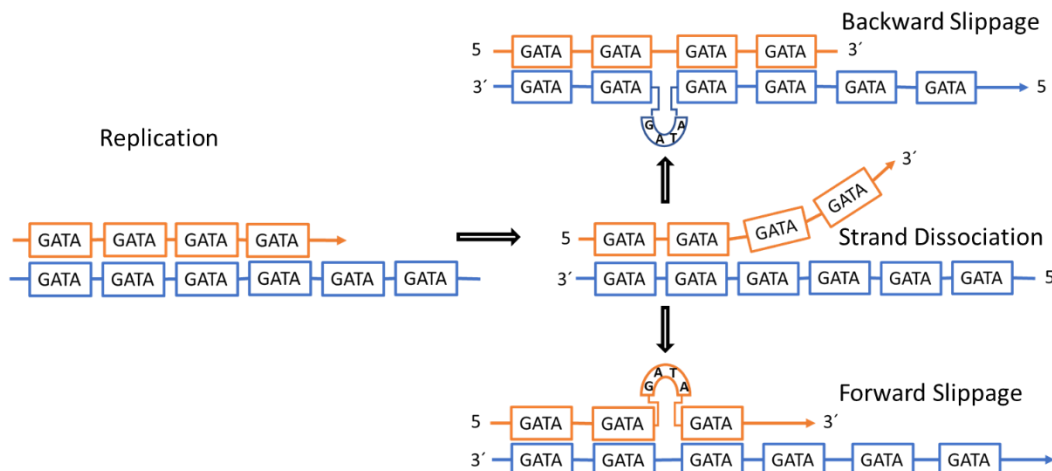


Figure 5. Schematic illustration of slipped strand model. During the replication process the DNA polymerase attaches itself to 3' end of the DNA and synthesizes identical complementary strands. Sometimes the strands dissociate, and the DNA polymerase complex becomes unpaired, and the replication stops. When the two strands pair up again, sometimes they mis-pair synthesizing a shorter copy or, more rarely, a larger copy. These faulty copies are called stutter, and produce stutter or shadow peaks that are, typically, one STR unit smaller or larger than the biological allele.

cycle, if no strand slippage occurs, we expect two stutter strands and six full-length alleles at the end of the third PCR cycle. In the absence of further stuttering, the proportion of the stutter product to full-sized allele would be 25%. Contrast this with a high-template scenario where, say, forty template copies are available for amplification. Here, even if stutter were to occur for one of the strands early in the cycling, it would unlikely that would occur for the other 39. Thus, in the next cycle we would have produced only 2 stuttered strands and 78 full-length strands. In the absence of any further slippage, we can expect that the relative stutter intensity would be 2.6% of the main allele.

Previous work by Malek¹³ demonstrated the probability that each strand stutters in a given PCR cycle is in the order of 1 in 1000. Thus, for a STR multiplex of, say, 10 heterozygous loci (i.e., 20 alleles), and a starting template count of one DNA copy, approximately 1 in 50 samples would be expected to exhibit stutter that is at least 100% of

the full-sized allele. This is a staggering result since in these circumstances it would be impossible to categorize the artifact as such given its relative intensity. The implication to casework is undeniable since recent forensic work has focused on developing laboratory pipelines engineered to detect the smallest level of DNA - i.e., in the single copy regime. In forensic science, single cell analysis has shown the ability to alleviate the challenges associated with DNA mixture analysis.¹⁴ In cases where single cell technology is not available, the interpretation of mixed DNA samples becomes challenging when there is a major: minor DNA mixture and the minor component is probative. Take, for example, a sexual assault vaginal swab where few sperm and many vaginal cells are mixed. Another case-type is of an equally mixed sample of multitudes of contributors like the handle of a firearm shared by multifarious individuals. In both cases having clear expectations of stutter intensities can aid in determining whether a peak in stutter position is stutter versus stutter and allele. That is, if the peak-in-question is in a stutter position and its peak height is greater than expectation, one can infer that peak contains signal from allele. If, however, the relative intensities of stutter changes with t_o , then modification to current heuristic practices may be required.

As such, this work seeks to explore the impact of template levels, t_o , on stutter ratios. Since previous findings demonstrate that sequence structure may have an effect on stutter propensity, the first hypothesis explored is that of per-allele dependance of stutter ratios. Once per-allele dependance is established, the average stutter ratio and stutter ratio distribution between low- and high-template samples are tested to determine if they are indistinguishable. If stutter ratios across t_o differ, the implication to casework is substantive

as it suggests template driven probabilistic models are desirable for bulk-mixture interpretation and for that of single-cell interpretation.

2. Material and Methods

2.1. Data Preparation

The data were received as a CSV file containing multiple single source samples (i.e., one person samples) consisting of signal from three main sources: noise, allele, and stutter artifact (as represented above in Figure 3). Single source samples are used to understand the behavior of stutter since signal from multiple contributors would be conflated by interfering genotypes. The genotype of single source samples was provided, which allowed for the sequestration of each peak into one of the aforementioned categories – i.e., allele, noise, stutter. Peaks falling in the electrophoretic positions indicating it was a true allele were classified as such. Any peak falling in a bin that was one STR repeat unit shorter than a known allele was designated as reverse stutter; conversely, peaks that fell in a bin position that was one repeat unit larger than a known allele were classified as forward stutter signal. The remaining non-zero peaks were classified as noise. Since the purpose of this study was to investigate relative stutter abundances, only loci with known genotype at least three repeat units apart were used. In addition, though noise is an inherent part of the signal and is known to confound the signal, previous work has shown that it is only observed in the relevant STR positions approximately 10% of the time and exhibits low peaks heights (i.e., 10s of RFUs). Thus, the confounding effects of noise on stutter or allele signal is expected to be negligible.

There were two main categories of sample type used in this study: those amplified using 0.25 ng (40 copies) of DNA and those amplified with 0.0313 ng (5 copies), which were respectively labeled as high- and low- copy samples. In total, the data consisted of 62 single source samples: 31 of high template mass and 31 of low template mass. All samples

were pre-processed to filter non-stutter PCR artifacts such as dye blobs, spikes, minus A, and pull up. Stutter ratios were calculated for all stutter-allele pairs, as described below.

2.2. Stutter

Since peak intensity/height is a proxy for the number of DNA amplicons synthesized during PCR, the relative peak heights positioned in STR locations within the electropherogram in the allelic position, a , and the reverse ($a-1$) or forward ($a+1$) positions are relevant. As such, the stutter ratio SR, becomes a reasonable means by which to explore the effects of template mass on relative stutter copy numbers.

$$SR_{a-1} = \frac{PH_{a-1}[RFU]}{PH_a[RFU]}$$

$$SR_{a+1} = \frac{PH_{a+1}[RFU]}{PH_a[RFU]}$$

Here, PH_a is the peak height of the allele given the known genotype, PH_{a-1} is the peak height of the peak located in reverse stutter position and PH_{a+1} is the peak height found in the forward stutter position. The stutter ratios were calculated for reverse stutter and forward stutter resulting in 434 ratios across both template mass classes (i.e., 0.25 ng and 0.0313 ng) across 5 STR loci.

2.2.1. Stutter ratio distributions

To analyze and compare stutter ratio distributions of low and high template samples, statistical analysis on a per-allele basis was performed. In almost all scientific disciplines, visual analysis plays a vital role in exploring, analyzing, and presenting scientific data.¹⁵ By comparing two histograms, differences in their stutter ratios

distribution patterns, modes, data ranges, and outliers can be recognized. Any quantitative differences were discussed.

To statistically evaluate the probability that the stutter distributions originate from a common source, a two-sample Komogorov-Smirnov test (KS test) was performed. Here, the statistic is the maximum absolute difference, d_{\max} , between two cumulative distribution functions, which is compared against the null distribution. If the probability that d_{\max} under the null hypothesis exceeds the experimental value was less than 0.005, the two distributions were classified as different. Notably the p-value of 0.005 is substantially smaller from the typical value of 0.05, and is the result of applying the Bonferroni Correction¹⁶, which is used to eliminate errors associated with multiple comparisons. Since 10 alleles across five loci are compared, p-value corrections are applied as decision threshold for a level of significance of 0.005.

2.2.2. Stutter Ratio analysis for mean values

To analyze and compare stutter ratio means between high and low template samples, visual and statistical analysis was again performed. Visual analysis was performed by plotting stutter ratio versus mass (0.25 ng, 0.0313 ng) as box plots for each allele. During this analysis, the median, interquartile ranges and general boxplot-shapes were compared. Statistical analysis of the mean was performed by a t-test to compare the experimental t-value against the t-distribution under the null. As previously described, Bonferroni Corrections were applied resulting in the p-value of threshold of 0.005.

3. Results and Discussion

3.1. Allele specific stutter ratios versus locus specific stutter ratios

In DNA analysis, it is important to understand and predict stutter behavior to improve the understanding and interpretation of the signal contained in a DNA profile.¹¹ Like its allelic cousin, stutter is a fluorescently tagged DNA molecule whose peaks cannot be easily distinguished from peaks that are of the same length as the allele or ‘wild-type’. This can have detrimental with serious downstream interpretation effects, since the forensic domain is entirely reliant upon the accurate deconvolution of genotypes using signal differences and peak intensities. If allele peaks are obfuscated by stutter peaks, genotype inferences become a challenge. Take, for example, the case of an evidence sample containing unknown quantities of DNA from an unknown number of contributors. In the forensic context, the genotypes must be inferred from the data after making assumptions about the possible number of contributors that comprise the signal. If the number of contributors is incorrectly inferred, the alleged genotypes that comprise the evidence will be incorrect. The foundation of the relevance has been explained by Slooten et al.¹⁷

This, therefore, suggests that stutter has the potential to significantly impact weights of evidence against suspects, particularly if it is the minor component to the DNA mixture profile that is probative. As a concrete example, consider a single locus exhibiting three peaks in the 12, 13, 16 allele positions. Here, the evidence may be explained as a one-person sample with genotype 13,16 (because 12 is taken to be stutter) or a 2-person sample wherein any one of the two donors could have genotypes 12,12; 12,13; 13,13, 13,16; 16,16 explaining the evidence. If the 12 is assumed to be stutter only, a suspect of 13,16 will be included as potential contributor to the evidence. Contrast that with all of the individuals

(any individual of a 12,12 or 12,13 or 13,13, etc.) who could not be excluded if a different stutter decision was applied.

Sutter peak height ratio thresholds of, typically, 15% of the allele have been used to mitigate some of these negative implications, but this heuristic does not solve the entire problem, since the 15% threshold is only effective if stable and applicable across all mass ranges.

In DNA analysis, it is therefore pivotal to differentiate a peak as an artifact or an allele before the assignment of the number of contributors in order to conduct reasonable genotype deconvolution. If this differentiation is not performed correctly, it can lead to the assignment of an additional contributor, causing n contributors to be assigned as $n + 1$. Similarly, if a true allele peak at a major contributor's stutter position is incorrectly filtered using too high a threshold, the true allele of a minor contributor can be lost to inference.

In addition, locus-specific stutter filters are typically applied, but allele-specific filter models can ultimately be more effective.¹⁸ In the GlobalFiler™ manual, for example, the stutter ratio filters are in the range of 4-16% and are set by adding the per-locus average stutter ratio with 3 times its standard deviation.¹⁹ But stutter ratios increase with allele length.¹⁸ This is consistent with the analysis performed in this work and demonstrated by the Figure 6, which is the box-plot of stutter ratios for five representative loci, differentiated by color, for samples amplified with normal levels - i.e. 0.25 ng or 40 copies. We see that in all cases the larger alleles, or the ones with more STRs, exhibit higher median stutter ratio values within a locus.

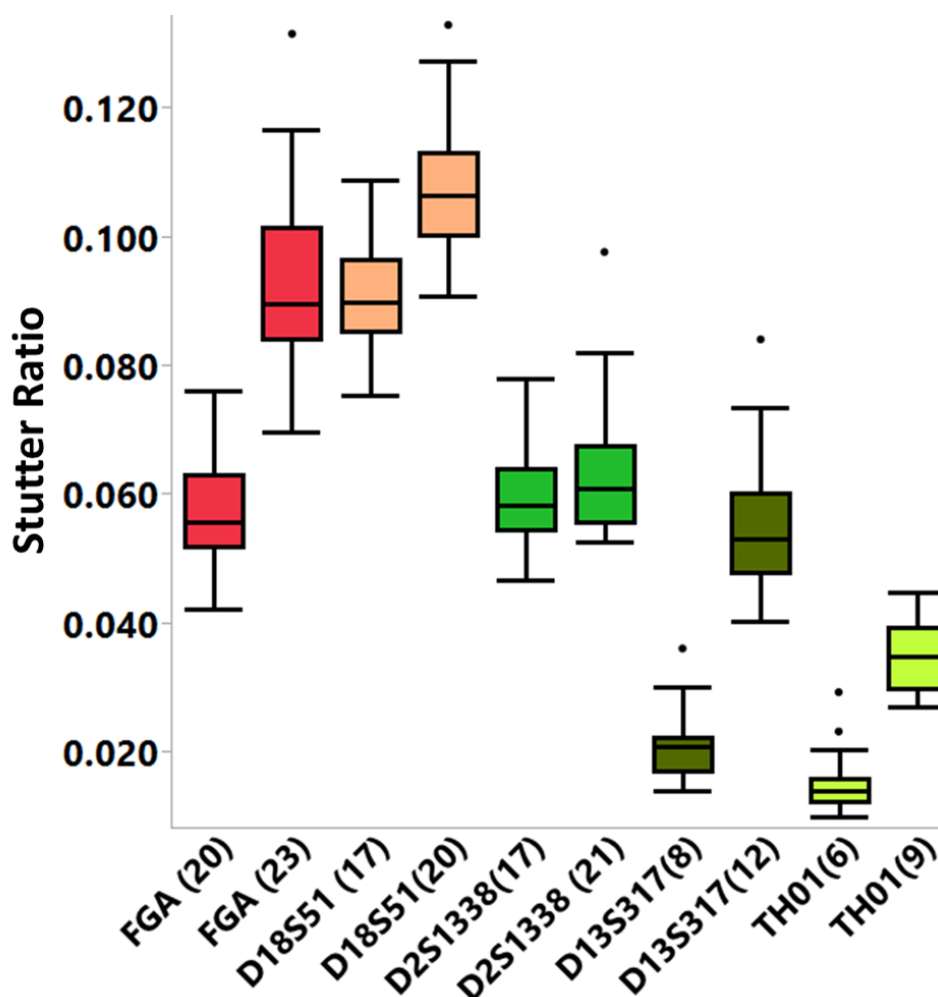


Figure 6. Box plots of high template stutter ratios against both alleles of five loci FGA (red), D18S51(yellow), D2S1338(green), D13S317(dark green), TH01(light green). Both alleles of each locus are expressed by same color box plots. Box plots show the median, interquartile range, and outer quartiles (black dots).

Visual analysis of the boxplot reveals that the larger allele within a locus has a higher median value and a larger range of stutter ratios. The stutter percentage difference between the alleles' median values is locus specific and ranges from 0.3 - 3.4%. Not only does the median value shift, but so does the entire distribution as the allele size increases. For example, FGA allele 20 exhibits a stutter ratio range from 4.2 - 7.6% while the larger sister allele 23 shows a range from 7 - 11.7%. Generally, stutter ratio increases with

increase in allele size.¹⁹

One key feature of the boxplot is that though an increase in stutter ratio is observed across both alleles within a locus, the magnitude of the change does not seem to trend with differences in STR length. At locus D2S1338(17, 21) the larger allele shows a minor increase in stutter ratio median value and distribution range with increase in allele length. This suggests that it is not the length of STR region alone that impacts the propensity for in-vitro strand slippage during PCR synthesis. For example, at loci D2S1338 (17,21), and D13S317(8,12) the difference in base pair length of both alleles is 4 but differences in median values and stutter ratio distribution range is noticeable. To further explore, Figure 7 is a scatter plot that presents the difference in median stutter ratios and differences in the STR length within the loci. The plot also depicts the results from a linear least square regression with R^2 0.073 indicating the majority of variation in median stutter values can be explained with features other than changes in STR length. Applying Pearson correlation, r : -0.27 shows a negative weak relationship between the difference in median values and the difference in STR size of both alleles, indicating additional factors influence the rate of stutter slippage.

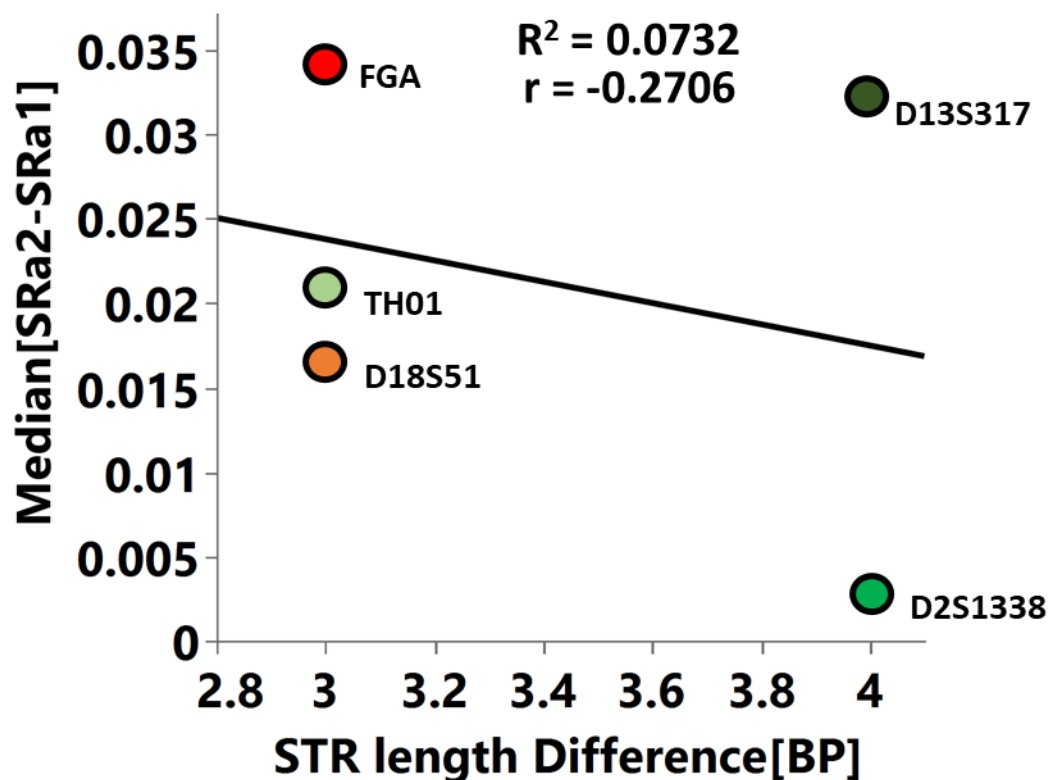


Figure 7. Scatter plot between difference in stutter ratio of both allele and difference in STR length of a STR locus.

As depicted in Figure 6, there is a relationship between relative stutter intensity and allele length, wherein stutter ratios increased with increases in the numbers of repeat units; however Figure 7 suggests there is another source of variation.²⁰⁻²² Some have suggested that the STR sequence plays a role. In particular, a relationship between stutter ratios and the longest uninterrupted sequence (LUS) has been proposed.^{11,21} If the allele consists of one repeat sequence, the stutter ratio distribution generally increases with the increase in allele length.^{19, 23} But if the allele contains several repeat sequences interrupted with a conserved or non-consensus segments, the stutter ratio distribution relationship can be better explained with LUS concept.¹¹ Other factors that may affect stutter formation include

weakly hybridized nucleotides, which cause higher rates of strand slippage, processivity differences between nucleases, and the uniformity and length of repeat motif.⁸

The STR repeat unit varies on the basis of repeat patterns which can be divided into three main categories: simple repeat, compound repeat, and complex repeat. Recall, the simple repeats consist of identical length and sequence, the compound repeats consist of two or more simple repeat units, and the complex repeat consists of multiple repeat blocks of variable unit length with variable intervening sequences. The loci interrogated in the study consist of all three: simple repeats (TH01, D18S51, D13S317); compound repeats (D2S1338); and complex repeats (FGA). To get a better understanding of their repeat motifs and relationship with their reverse ratios distribution they are each described, in turn.

3.1.1. TH01

TH01(6,9) is a tetrameric short tandem simple repeat marker with an [AATG]_n repeat sequence.⁴ It is one of the most widely used markers in forensic casework. TH01(6,9) consists of smaller allele [AATG]₆ and larger [AATG]₉, so, with the increase of allele number, an increase in median value and upward increase in stutter ratio is observed. The observed stutter percentages are consistent with the manufacturer's (Identifiler™ Plus Amplification Kit User Guide) observations and other scientific literature.^{11, 19} Though not sampled in this study a common incomplete STR (9.3) [AATG]₆ATG[AATG]₃ is known.⁴ The seventh repeat sequence has a one nucleotide missing. Interestingly when the authors compared 9.3's stutter ratio with that of 6, it was indistinguishable, once again suggesting sequence structure can have drastic effects on stutter production.¹¹ In addition, TH01 has a higher adenine to thymine ratio 3/4 which

increases the chance of stutter product formation, since A-T only has two hydrogen bonds, while G-C have three.¹¹ Given the observed impacts of allele-type on stutter ratios, and the supporting literature, statistical analysis is performed on a per-allele basis.

3.1.2. D18S51

D18S51(17,21) is a tetrameric short tandem simple repeat marker with simple [AGAA]_n sequence. The repeat units have relatively high levels of A-T, which have a higher tendency to produce stutter products. D18S51(17,21) consists of a smaller allele 17, [AGAA]₁₇, and the larger allele 21, [AGAA]₂₁. As the number of repeat units grow larger, higher stutters ratios ranges 7.5% to 12.7% were observed in Figure 6. The stutter ratio distribution ranges were consistent with the manufacturer. Once again this locus demonstrates allele specific stutter filters, could reduce the chance of over-, or under-filtration of alleles.¹¹

3.1.3. FGA

FGA is a complex short tandem repeat unit with multiple sequences making up the total STR: [TTTC]₃TTTT TTCT[CTTT]_nCTCC[TTCC]₂.⁴ Its repeating pattern is highly polymorphic, providing it an advantage of high discrimination power. This is the reason it is commonly used in laboratories and is a part of most commercially available kits.²⁴⁻²⁵ FGA has two allelic groups based on their size: alleles 16-34.2 and 42.2-51.2.^{19, 21} Specifically related to this study, FGA 20 has a repeat sequence [TTTC]₃TTTT TTCT[CTTT]₁₂CTCC[TTCC]₂, and allele 23 has structure [TTTC]₃TTTT TTCT[CTTT]₁₅CTCC[TTCC]₂. In both, repeat sequence [CTTT]_n shows 12, and 15 repeat units for STR 20 and 23 respectively.⁴ According to the LUS model, an allele with multiple repeat sequences stutters on the base of its LUS. FGA allele 20 has 12 LUS units. Visual

analysis of the Figure 6 reveals FGA allele 20 with median stutter ratio (5.6%) tracks well with D13S317 (simple repeat) allele 12's median value of 5.3%.^{19,23} Similarly, FGA allele 20 stutter ratio distribution range 3.4% (4.2-7.6%) compared with D13S317 allele 12 range 3.3% (4-7.4%) is almost the same. In addition, these observations are consistent with previous studies^{19,23}.

Similarly, FGA 23 with LUS [CTTT]₁₅ shows a median value of 9% while D18S51 allele 17 with sequence [AGAA]₁₇ resulted in the same median value (9%).

3.1.4. D2S1338

D2S1338 [TGCC]_n[TTCC]_n is a compound motif consisting of two repeat units.⁴ In Figure 6 its smaller allele is 17 and the larger one is allele 21. There is a difference of four STR repeat units, but, interestingly, exhibit similar median values, with only a difference of 0.3%. This is because of the motif structure of D2S1338 [TGCC]_n[TTCC]_n. For the smaller allele 17, the repeat structure is plausibly [TGCC]₆[TTCC]₁₁. TTCC is the longest uninterrupted stretch and is likely to act as the core repeat unit and produce stutter products accordingly. TGCC will also likely produce lower-level stutters but to a much smaller degree. This is, indeed, borne out by the data that shows D2S1338 17's median value is close to D13S317 allele 12's median, once again supporting the per-allele statistical analysis to mitigate confounding sequence-dependencies in the following section.

3.2. Stutter Ratio between high and low template samples

Though previous literature shows that stutter ratio behavior changes in the low template mass regime, and that low template stutter ratios are higher, these studies do not address the question on a per allele basis. In response, this work aims to address this gap

by evaluating if low- and high- template relative stutter abundances are of distinct distributions. By acknowledging the literature, and the dependency of stutter on sequence, we re-evaluate stutter distributions on a per-allele basis for both the reverse and forward stutter types. For exposition we present, in detail, the results obtained from four loci: FGA; TH01; D2S1338; D18S51. Notably, all data are from samples whose known STR alleles are at least three repeat units apart in order to avoid signal conflation between allele and stutter.

3.2.1. Reverse Stutter

3.2.1.1. FGA (20)

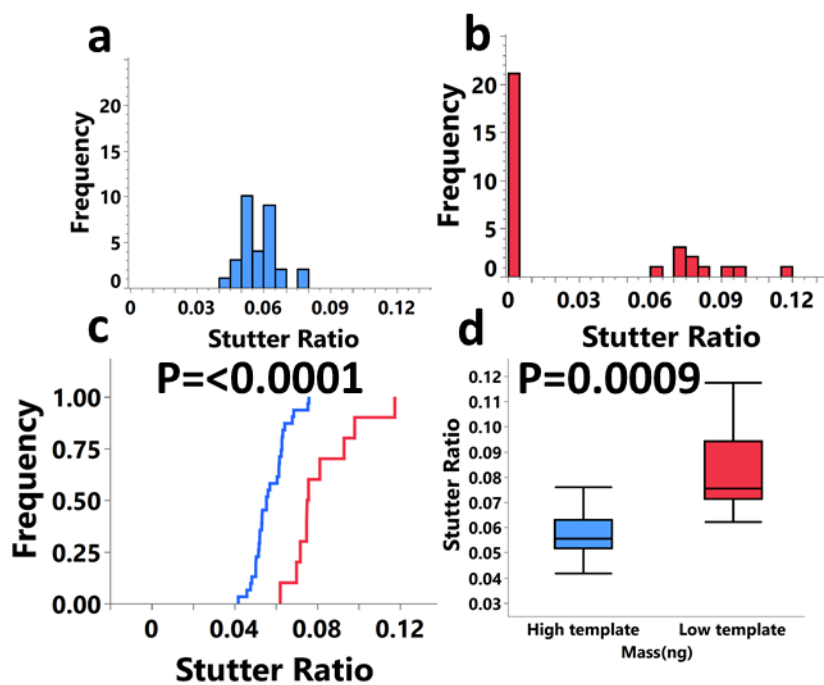


Figure 8. Reverse Stutter: (a) A histogram (blue) of allele 20(FGA) with base pair size 225 bp for the high template samples. (b) A histogram (red) of allele 20(FGA) for the low template samples. (c) A KS- test plot showing cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.835$ and $p < 0.0001$. (d) A box plot 20(FGA) of all non-zero stutter ratios which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value 4.56 and p-value 0.0009.

The stutter ratio data distribution for high template samples ranges from 4% to 7%, which is in-line with the values reported by the manufacturer (Identifiler™ Plus Amplification Kit). In contrast, for the low template stutter ratios the stutter ratios ranged from 6% to 10%; that is, the distribution shifted to the right. As low template sample stutter ratios are larger, they can make DNA profile interpretation more complicated

since a forensic DNA mixture is the result of a combination of any number of contributor's DNA mixed in any proportion.

A KS-test was performed to determine if reverse stutter distributions of low and high template data are insignificantly different for the 20 alleles. Figure 8 (c) illustrates the representative plot of stutter ratio distributions of low and high template sample data. In the KS-test the cumulative frequencies of stutter ratios were compared between low and high template data distributions, where $d = 0.835$, and $p < 0.0001$. The resultant p-value is less than the critical value 0.005, suggesting the high and low template reverse stutter ratio distributions are significantly different.

Figure 8 (d) is a box plot of stutter ratio versus mass (0.25 ng, 0.0313 ng) of allele 20 of locus FGA. Firstly, by comparing the median values of the two distributions, one can see that the high template stutter ratio has a smaller median (5.6%) then the low template class (7.5%). The increase in stutter ratio with a decrease in template mass shows stutter ratio percentages move upward in its central tendency. When the stutter ratio value of high template samples are compared with the estimated median value from commercially available amplification kits: Identifiler™ Plus Amplification Kit²³ the two are consistent at ca. 5.6%. When these values are compared with low template sample median values, we observe a difference around 2.0%. Secondly, by comparing the interquartile range of two template samples it shows that the high template samples have a smaller interquartile range (1.2%) while low template have a higher one (2.3%), meaning for high template stutter ratio data concentrates more around the median then their low template counterparts.

The average stutter ratio for low template samples average 8.2% which is consistent with the low- template values of Seo et al.'s who amplified 0.03 ng DNA for 32 and 34

PCR cycles and obtained stutter ratio averages of 11% and 9.8%, respectively.²⁶ So, by comparing the stutter ratio distributions, median and mean values of two template samples it can be concluded, for FGA allele 20, low template samples enhance the stutter percentage in their central tendency and overall distribution. To confirm this statistically, a two tailed t-test was performed. When comparing the of p-value 0.0009 against the Bonferroni corrected critical value of 0.005, the null hypothesis is rejected, supporting the claim that the two means are different.

3.2.1.2. FGA (23)

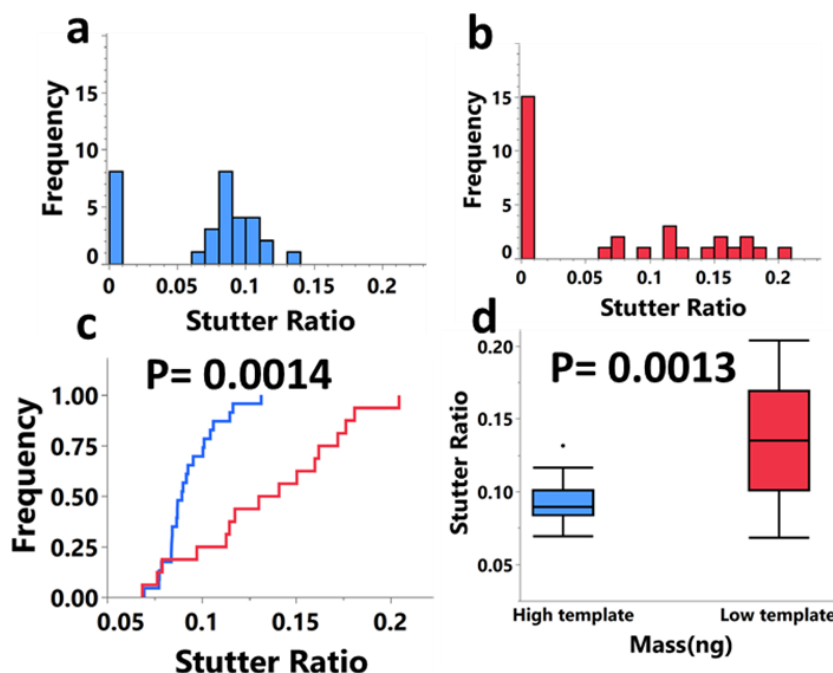


Figure 9. Reverse Stutter: (a) A histogram (blue) of allele 23(FGA) with base pair size 237 bp for the high template samples. (b) A histogram (red) of allele 23(FGA) for the low template samples. (c) A KS-test plot of cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.620$ and $p = 0.0014$. (d) A box plot 23(FGA) of all non-zero stutter ratios, which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value of 3.8 and p-value 0.0013.

FGA 23 [TTTC]₃TTTT TTCT[CTTT]₁₅CTCC[TTCC]₂ is a complex motif with [CTTT]₁₅ repeat unit. For Figure 9 (a), the first bin on the left side shows the complete absence of stutter for 8 of the 31 stutter positions tested. This is an unusual finding since high template results typically render stutter intensities greater than zero relative fluorescence unit (RFU). Stutters are produced as by-product during the PCR reactions.

The high template stutter ratio distribution's range is 7.5-12.5%. When this range is compared with the manufacturer's range of 4.6-9.5%, we see they are consistent. For

low template samples the stutter percentage is between 5-22.5%, which is, arguably different from the high template results.

To test that assertion, a KS-test was performed to determine if stutter ratio distributions of low and high template data are insignificantly different for allele 23 FGA. Figure 9 (c) illustrates the representative plot of stutter ratio cumulative distributions of low and high template sample data. The cumulative frequencies showed a maximum difference of $d = 0.62$ and $p = 0.0014$, suggesting the distributions are not alike.

Lastly, Figure 9 (d) is a box plot of stutter ratio versus mass (0.25 ng, 0.0313 ng) of a representative allele 17 of locus D18S51. By comparing the median values high template stutter ratio has a lower median at about 0.09 while low template has 0.11.

Unlike the median, the average stutter ratio values for high and low template samples seem different and are 9.3%, and 13.4% which is not surprising given the means, in general, are more sensitive to outliers. A t-test provides a t-statistic of 3.8, suggesting there are, indeed, differences in the means p-value: 0.0013.

Similarly, visual and statistical analysis were performed per allele for the other four loci to compare if there are insignificant differences between high and low template stutter ratios distributions and average values (see Appendix 1). The summary of results is presented in the following table.

Table 1. A summary of p-value obtained from KS-test and t-test, the ranges and the proportions of zero-stutter values for five loci tested across the high template and low template category.

Locus/allele	Template-mass class	Percent zero stutter	Range	p-value KS-test	p-value t-test
FGA (20)	High	-	4 – 8%	<.0001	0.0009
	Low	67.7%	6 – 12%		
FGA (23)	High	25.8%	6 – 14%	0.0014	0.0013
	Low	48.4%	6 – 21%		
TH01 (6)	High	-	0.75 – 3%	<.0001	0.0035
	Low	61.3%	1.5 – 7.8%		
TH01 (9)	High	74.2%	2.5 – 4.5%	0.6749	0.3052
	Low	48.4%	2.3 – 5.8%		
D2S1338(17)	High	-	4.5 – 8%	0.0035	0.0012
	Low	16.1%	3 – 11%		
D2S1338(21)	High	-	5 – 10%	0.0006	0.0006
	Low	38.7%	4.5 – 14%		
D18S51(17)	High	-	7 – 11%	0.0008	0.0076
	Low	45.2%	7 – 21%		
D18S51(20)	High	-	9 – 14%	<.0001	0.0013
	Low	19.4%	9 – 17%		
D13S317(8)	High	3.2%	1.3 – 3.8%	0.0076	0.0525
	Low	48.4%	1.5 – 4.3%		
D13S317(12)	High	-	4 – 8.5%	0.2822	0.1347
	Low	29%	4 – 8%		

D18S51(20) low template stutter ratio has an outlier in 26 - 27% range.

D13S317(8) low template stutter ratio has an outlier in 7.3 - 7.5% range.

D13S317(12) low template stutter ratio has two outliers in 10 – 12.5% range.

Taken together, Figures 8 and 9, and Table 1 demonstrate that the low-template mass category cannot be classified as originating from the same distribution, which is

pertinent since in many cases high-template stutter ratio thresholds, calculated by adding three times the standard deviation to the mean, are used as the filtering rule for casework. This work demonstrates, however, that higher filters as per Table 2 for low-template sample are ought to be considered for operations. Alternatively, continuous models that model stutter's relative abundance with respect to some proxy for template mass is a viable option.

Table 2. A summary of SRT obtained from average stutter ratios, standard deviation values for low and high template samples per- allele across five loci.

Locus/allele	Template-mass category	SR	STDEV SR	SRT
FGA (20)	High	0.0574	0.0082	8.2%
	Low	0.0820	0.0164	13.1%
FGA (23)	High	0.0929	0.0145	13.6%
	Low	0.1337	0.0409	25.7%
TH01 (6)	High	0.1461	0.0039	2.6%
	Low	0.0321	0.0164	8.1%
TH01 (9)	High	0.0350	0.0058	5.2%
	Low	0.0386	0.0110	7.2%
D2S1338(17)	High	0.0598	0.0074	8.2%
	Low	0.0737	0.0187	13%
D2S1338(21)	High	0.0630	0.0106	9.5%
	Low	0.0888	0.0267	16.9%
D13S317 (8)	High	0.0206	0.0051	3.6%
	Low	0.0282	0.0141	7.1%
D13S317(12)	High	0.0540	0.0096	8.3%
	Low	0.0607	0.0188	11.7%
D18S51(17)	High	0.0911	0.0087	11.7%
	Low	0.1180	0.0360	22.6%
D18S51(20)	High	0.1076	0.0102	13.8%
	Low	0.1341	0.0335	23.5%

3.2.2. Forward Stutters:

Like reverse stutter, during the amplification process, forward stutters are produced as introduced in Figure 5.²² In DNA analysis, forward stutter generally has less interference with mixture interpretation as compared to backward stutters due to a relatively low level of stutters, though it can still impact casework when the probative component to the mixture is minor.

Forward stutters were observed at two loci D18S51(17, 21) and D2S1338(17,21) of the data. Like reverse stutters, visual and statistical analysis was performed to evaluate if there is an insignificant difference in high and low template stutter ratios distributions and average values. The results obtained are summarized in the following table.

Table 3. A summary of p-value obtained from KS-test and t-test, the ranges, and the proportions of zero-stutter values for 2 loci tested across high and low template categories of forward stutters.

Locus/allele	Template-mass class	Percent-zero stutters	Range	p-value KS-test	p-value t-test
D18S51(17)	High	29%	1 – 3%	<.0001	0.4830
	Low	87.1%	4.3 – 9%		
D18S51(20)	High	54.8%	0.5 – 7%	0.1466	0.29
	Low	83.9%	2.3 – 4.3%		
D2S1338(17)	High	96.8%	0.4 – 0.6%	1.0	0.2108
	Low	93.6%	2.2 – 3.6%		
D2S1338(21)	High	90.3%	0.3 – 2.6%	1.0	0.9631
	Low	96.8%	2.9 – 3%		

4. Conclusion

Forensic mixture deconvolution requires a strategy that takes into account the pre-processing steps. There are many instances where stutter ratios are modelled as a continuous variable, with probability p that it occurs given the main allele exhibits height, h . In cases where non-continuous methods are employed, stutter filters are used. In the current study, we have explored the impact of template levels, t_o , on stutter ratios. The findings demonstrate that in a low-template regime, in the absence of fully continuous models, the relative abundance of stutter increases in the low-template regime; thus, increasing the stutter ratio threshold is one viable option. On the basis of this observation, it can also be concluded that it is important to efficiently characterize low-template samples' stutter ratios and establish clear guidelines for their DNA mixture interpretation. In addition, acknowledging previous literature and findings within this work, it is evident that allele-specific thresholds reduce the chance of incorrect interpretation of stutter peaks. Consequently, allele-specific thresholds should be preferred over locus-specific thresholds.

Overall, this study has confirmed previous work²⁷ that stutter ratios differ across the samples' template mass; hence, its implication to casework is substantive. Therefore, it is suggested that template-driven probabilistic models should be applied in bulk-mixture and single-cell interpretation.

Appendix 1

A representation of simple motif: TH01 allele 6 high and low template stutter ratios to compare if there is insignificant difference between high and low template stutter ratios distributions and average values.

TH01 (6)

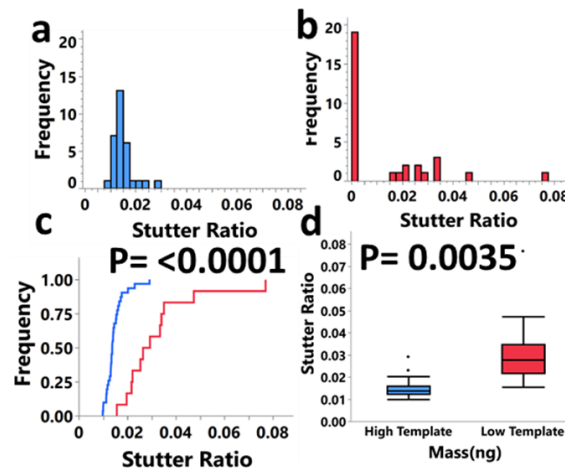


Figure 1. Reverse Stutters: (a) A histogram (blue) of allele 6(TH01) with base pair size 168 bp for the high template samples. (b) A histogram (red) of allele 6(TH01) for the low template samples. (c) A KS-test plot showing cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.820$, and $p = <0.0001$. (d) A box plot 6(TH01) of all non-zero stutter ratios, which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value 3.7, and p-value 0.0035.

A representation of simple motif: TH01 allele 9 high and low template stutter ratios to compare if there is insignificant difference between high and low template stutter ratios distributions and average values.

TH01 (9)

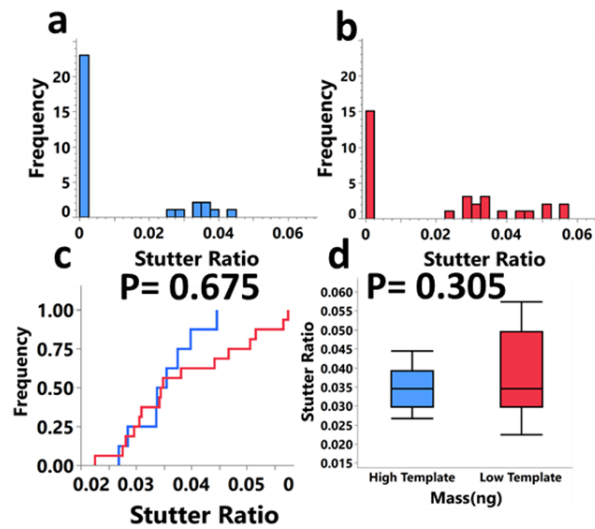


Figure 2. Reverse Stutters: (a) A histogram (blue) of allele 9(TH01) with base pair size 180 bp for the high template samples. (b) A histogram (red) of allele 9(TH01) for the low template samples. (c) A KS-test plot showing cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.313$ and $p = 0.675$ (d) A box plot 9(TH01) of all non- zero stutter ratios, which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value 1.1, and p-value 0.3052.

A representation of compound motif: D2S1338 allele 17 high and low template stutter ratios to compare if there is insignificant difference between high and low template stutter ratios distributions and average values.

D2S1338 (17)

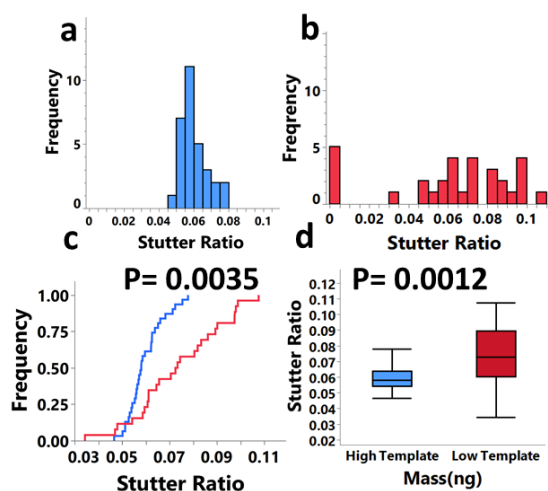


Figure 3. Reverse Stutters: (a) A histogram (blue) of allele17(D2S1338) with base pair size 313 bp for the high template samples. (b) A histogram (red) of allele 17(D2S1338) for the low template samples. (c) A KS-test plot of the cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.474$, and $p = 0.0035$. (d) A box plot 17(D2S1338) of all non-zero stutter ratios, which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test gives a t-value 3.5, and p-value 0.0012

A representation of compound motif: D2S1338 allele 21 high and low template stutter ratios to compare if there is insignificant difference between high and low template stutter ratios distributions and average values.

D2S1338 (21)

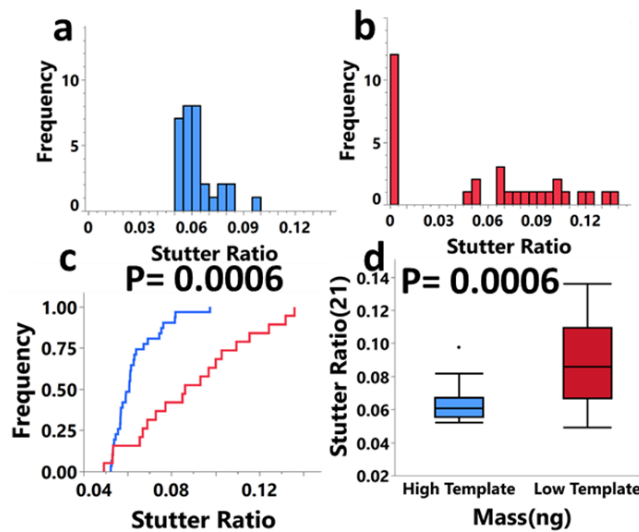


Figure 4. Reverse Stutters: (a) A histogram (blue) of allele 21(D2S1338) with base pair size 329 bp for the high template samples. (b) A histogram (red) of allele 21(D2S1338) for the low template samples. Figure (c) A KS-test plot showing cumulative frequency versus stutter ratios of high (blue) and low (red) template mass, where $d = 0.584$, and $p = 0.0006$. Figure (d) A box plot 21(D2S1338) of all non-zero stutter ratios which compares the means of high (blue box) and low (red box) template stutter ratios. A t-test was performed to compare if the means of the two stutter ratios are insignificantly different. The t-test has a t-value 4.0, and p-value 0.0006

Bibliography

1. Riman, S.; Iyer, H.; Borsuk, L. A.; Vallone, P. M., Understanding the behavior of stutter through the sequencing of STR alleles. *Forensic Science International: Genetics Supplement Series* **2019**, 7 (1), 115-116.
2. Budowle, B.; Eisenberg, A. J.; van Daal, A., Validity of Low Copy Number Typing and Applications to Forensic Science. *Croatian Medical Journal* **2009**, 50 (3), 207-217.
3. Butler, J. M., *Forensic DNA Typing*. Second Edition ed.; Elsevier Academic Press: Burlington, MA, USA, 2005.
4. technology., N. I. o. S. a. Short Tandem Repeat DNA Internet DataBase.
5. Butler, J., *Advanced Topics in Forensic DNA Typing: Methodology*. 1st Edition ed.; Academic Press 2011: San Diego, July 2011; p 708.
6. Evans, M. F., The polymerase chain reaction and pathology practice. *Diagnostic Histopathology* **2009**, 15 (7), 344-356.
7. Ruijter, J. M.; Ramakers, C.; Hoogaars, W. M. H.; Karlen, Y.; Bakker, O.; van den Hoff, M. J. B.; Moorman, A. F. M., Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res.* **2009**, 37 (6), e45-e45.
8. Butler, J. M., *Advanced Topics in Forensic DNA Typing: Methodology*. Elsevier Science & Technology: San Diego, UNITED STATES, 2011.
9. Tao, R. Y.; Wang, S. Y.; Zhang, J. S.; Zhang, J. Y.; Yang, Z. H.; Sheng, X.; Hou, Y. P.; Zhang, S. H.; Li, C. T., Separation/extraction, detection, and interpretation of DNA mixtures in forensic science (review). *Int. J. Legal Med.* **2018**, 132 (5), 1247-1261.
10. Hauge, X. Y.; Litt, M., A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Human Molecular Genetics* **1993**, 2 (4), 411-415.
11. Brookes, C.; Bright, J. A.; Harbison, S.; Buckleton, J., Characterising stutter in forensic STR multiplexes. *Forensic science international. Genetics* **2012**, 6 (1), 58-63.]
12. Bright, J. A.; Buckleton, J. S.; Taylor, D.; Fernando, M.; Curran, J. M., Modeling forward stutter: toward increased objectivity in forensic DNA interpretation. *Electrophoresis* **2014**, 35 (21-22), 3152-3157.

13. Malik, L. STABILIZING INFORMATION CONTENT IN DNA EVIDENCE TO IMPROVE LAB-TO-LAB INTERFERENCE. Rutgers, The state University of New Jersey, Camden, New Jersey, 2019.
14. Watkins, D. R. L.; Myers, D.; Xavier, H. E.; Marciano, M. A., Revisiting single cell analysis in forensic science. *Scientific Reports* **2021**, *11* (1), 7054.
15. Kehrer, J.; Hauser, H., Visualization and visual analysis of multifaceted scientific data: a survey. *IEEE transactions on visualization and computer graphics* **2013**, *19* (3), 495-513.
16. Armstrong, R. A., When to use the Bonferroni correction. *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)* **2014**, *34* (5), 502-8.
17. Slooten, K.; Caliebe, A., Contributors are a nuisance (parameter) for DNA mixture evidence evaluation. *Forensic Science International: Genetics* **2018**, *37*, 116-125.
18. Kalafut, T.; Schuerman, C.; Sutton, J.; Faris, T.; Armogida, L.; Bright, J.-A.; Buckleton, J.; Taylor, D., Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation. *Forensic Science International: Genetics* **2018**, *35*, 50-56.
19. Scientific, T., GlobalFiler™ PCR Amplification Kit User Guide. Life Technologies Corporation: Carlsbad, CA 92008 USA, 07 July 2016.
20. Klintschar, M.; Wiegand, P., Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. *Forensic Sci.Int.* **2003**, *135* (2), 163-166.
21. Lazaruk, K.; Wallin, J.; Holt, C.; Nguyen, T.; Walsh, P. S., Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing. *Forensic Sci.Int.* **2001**, *119* (1), 1-10.
22. Walsh, P. S.; Fildes, N. J.; Reynolds, R., Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* **1996**, *24* (14), 2807-2812.
23. Technologies, A. B. L., AmpFISTR™ Identifiler™ Plus PCR Amplification Kit User Guide. Thermo Fisher Scientific: Warrington Cheshire United Kingdom, 19 September 2018; Vol. H.
24. Ariel, T.; Dell'Ariccia-Carmon, A.; Pasternak, Z.; Raziell, A., Population-dependent migration shift caused by sequence variation at the alpha fibrinogen (FGA) short tandem repeat (STR) locus. *Forensic Sci.Int.* **2020**, *314*, 110395.

25. Gill, P.; d'Aloja, E.; Andersen, J.; Dupuy, B.; Jangblad, M.; Johnsson, V.; Kloosterman, A.; Kratzer, A.; Lareu, M.; Meldegaard, M., Report of the European DNA profiling group (EDNAP): an investigation of the complex STR loci D21S11 and HUMFIBRA (FGA). *Forensic Sci.Int.* **1997**, *86* (1-2), 25-33.
26. Seo, S. B.; Ge, J.; King, J. L.; Budowle, B., Reduction of stutter ratios in short tandem repeat loci typing of low copy number DNA samples. *Forensic science international. Genetics* **2014**, *8* (1), 213-8.
27. Alfonse, L. E.; Garrett, A. D.; Lun, D. S.; Duffy, K. R.; Grgicak, C. M., A large-scale dataset of single and mixed-source short tandem repeat profiles to inform human identification strategies: PROVEDIt. *Forensic Science International: Genetics* **2018**, *32*, 62-70.