

Open Access

A Simple Method of Generating Complex DNA Profiles Utilizing *Alu*-based Markers with Applications in Forensics, Paternity, Genetic Mapping, Population Studies and Ancestry

David H Kass*

Department of Biology, 441 Mark Jefferson Ypsilanti, MI 48197 USA

Abstract

As DNA profiling systems become more complex, advancements to a relatively simple technique are presented that promote greater accessibility and usefulness for a variety of applications. In contrast, other simple tools commonly used to teach students about forensics and human populations have notable drawbacks. Two *Alu*tetraplex systems, utilizing four *Alu* presence/absence variants in a single reaction were therefore developed to provide a simple methodology to generate complex profiles. A third *Alu*tetraplex system is presented, escalating the number of possible genotypes to 531,441, with all alleles of the 12 dimorphic markers being relatively common. Reproducible results were attained even with the use of crude DNA preparations stored frozen for several years with multiple freeze thaws. The incorporation of GelRed DNA stain instead of the highly toxic ethidium bromide promotes greater accessibility, particularly in a classroom setting. This study demonstrates the effectiveness of this profiling system as a simple but informative methodology to analyze paternity, genetic mapping of human traits, and provides data to illustrate its potential in assessing ancestry or geographic origins of an individual.

Keywords: *Alu* dimorphism; *Alu*tetraplex; Ancestry; Forensics; Positional cloning; DNA markers; DNA profiles

Introduction

The D1S80 variable number tandem repeat (VNTR) and the PV92 dimorphic *Alu* element are commonly incorporated DNA profiling tools for educational purposes that are useful in teaching students about forensics and human population studies [1]. However, these simple, low-cost profiling tools have notable drawbacks. The resolution limitation yields difficulty in assessing alleles of the D1S80 locus, particularly from agarose gels; whereas the PV92 locus only



Figure 1: Schematic diagram of an *Alu* integration into a chromosome yielding a dimorphic presence/absence marker which can be analyzed by PCR using forward (F) and reverse (R) primers that flank the integration. By agarose gel electrophoresis the three possible genotypes could be determined as the *Alu* integration yields a fragment approximately 300 bp larger than the absence (ancestral) allele.

consists of two alleles (presence or absence of the Alu) or three possible genotypes and therefore is not overly informative. PV92 is an example of an intermediate-frequency (IF) Alu, whereby both the presence and absence forms, also known as a dimorphism, are common in human populations. These types of variants result from a relatively recent integration event of the Alu retrotransposon (Figure 1), and therefore not fixed in the human genome. Approximately 1,200 of the over one million Alu elements within the human genome are dimorphic [2]. The presence and absence alleles can be distinguished by the polymerase chain reaction (PCR) utilizing primers that flank the element (Figure 1). A number of IF Alu variants have been identified [3,4] and therefore, combining several of these markers in a single reaction exponentially increases the possible number of genotypes without increasing the amount of reactions. A major obstacle in generating these profiles by multiplex PCR, however, is the formation of heteroduplexes, as very similar Alu elements are found in each presence variant, and limits the capability of standard Taq DNA polymerase with a variety of conditions [5-7]. The Failsafe system (Epicentre Biotechnologies, Madison, WI) designed for difficult or multiplex PCR has been demonstrated to work well with up to four IF Alu loci in a single reaction, referred to as an Alutetraplex, with the primary adjustment being the modification of primer concentrations [6,7].

*Corresponding author: David H Kass, Department of Biology, 441 Mark Jefferson Ypsilanti, MI 48197 USA, Tel: 734-487-2388; Fax: 734-487-9235; E-mail: dkass@emich.edu

Received July 17, 2013; Accepted September 23, 2013; Published September 27, 2013

Citation: Kass DH (2013) A Simple Method of Generating Complex DNA Profiles Utilizing *Alu*-based Markers with Applications in Forensics, Paternity, Genetic Mapping, Population Studies and Ancestry. J Forensic Res 4: 200. doi:10.4172/2157-7145.1000200

Copyright: © 2013 Kass DH. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Alutetraplexes offer a relatively low-cost technology that is simple to use, requires only standard laboratory equipment, and yet provides relatively complex DNA profiles. This technique provides individuals, particularly students, the opportunity to perform forensic analyses by utilizing their DNA samples and matching it to an unknown, a sample selected by the instructor [7]. Along with a constructed database of anonymous submissions [8], it has also allowed the students to determine the proportion of individuals with the same genotypes, as well as perform a variety of population analyses such as genetic distances and assessment of Hardy-Weinberg equilibrium. The two developed Alutetraplexes [7] generate 6,561 possible genotypes (34×34), far exceeding the 435 possible D1S80 genotypes. Additionally, one IF Alu locus has been found to consist of four alleles [9] providing a unique tool in studying human populations, particularly considering the evolutionary order of the generation of the alleles can be determined. Along with forensic and population analyses, this simple technique demonstrates how DNA markers can be used in paternity [7].

Polymorphic DNA markers of various types have been used in positional cloning for the purpose of mapping the chromosomal position of a gene associated with a trait with the objective of eventually identifying that gene [10]. A variable number tandem repeat (VNTR) or a microsatellite marker has greater potential than biallelic markers such as restriction fragment length polymorphisms (RFLPs), by having a greater possibility of a heterozygous genotype for the marker in the individual with the trait. However, large data sets of biallelic markers such as a map of 1.42 million single nucleotide polymorphisms (SNPs) distributed throughout the human genome [11] provides a highly useful tool to aid in identifying biomedically important genes. An Alutetraplex consists of four biallelic markers in a single reaction. Therefore, if three tetraplex reactions were available, the potential of finding a useful marker (heterozygous genotype) increases, giving the student an opportunity to perform a mapping experiment and evaluate significance by generating Lod scores [12]. This type of study would involve a non-health risk condition such as the autosomal dominant compelling helio-ophthalmic outburst (ACHOO) syndrome [13]. This syndrome is found in roughly 30% of the population and involves sneezing when initially exposed to bright light, also known as a photic sneeze reflex [14].

To enhance the *Alu*tetraplex methodology for various applications, a third tetraplex of IF *Alu* variants was developed in this study. By escalating the number of possible genotypes that can be determined, this simple technique becomes a more prominent tool in forensic analysis. Alternative techniques incorporating as many as 30-32 *Alu* dimorphisms, in studies of specific populations, have demonstrated high levels of discriminatory power in the range of 3.7×10^{-13} [15] to 5.53×10^{-14} [16] probability of a genotype match. The use of a third set of markers also increases the probability of addressing paternity. For example, if the mother is heterozygous for all four markers in one tetraplex, then the results are uninformative, but more markers will dramatically reduce this possibility and further aid in excluding individuals as the father. Additionally, this study presents the use of *Alu*-based tetraplexes as a tool for demonstrating genetic mapping of human traits.

The practical application of this method is supported by the use of crude rapid-extraction DNA preparations from buccal cells that have been stored at -20°C for several years with multiple freeze thaws. Additionally, the use of GelRed (Biotium, Hayward, CA) for analyzing tetraplex gels is presented as a safe DNA stain alternative to the highly toxic standard of ethidium bromide, making this technique more accessible, especially for classroom use. Overall, by generating a third tetraplex, the number of markers increases among different chromosomal sites, and markedly enhances the methodology as a simple, rapid, and low-cost tool for paternity, population, genetic mapping and forensic analyses, and may provide information regarding ancestry or geographic origins.

Materials and Methods

DNA samples

Crude, preparations of DNA were obtained using the Epicentre BuccalAmp DNA Extraction Kit protocol (www.epibio.com) from cheek swabs. Approval was obtained through the institutional human subjects review panel.

Genetic mapping

Two families, in which the father and two children had the ACHOO syndrome, but not the mother, volunteered their DNA samples for this study. The *Alu*tetraplex PCR conditions and primers are provided in a previous study [7]. *Alu* variants that are heterozygous in the father and homozygous in the mother would be useful in assessing linkage. Lod scores to statistically measure linkage were determined using the formula $\text{Log}_{10} [(1/2 (1-\theta)^{\text{nr}} \times \theta^{\text{r}})/(0.5)^{\text{n+r}} + (1/2 (1-\theta)^{\text{nr}} \times \theta^{\text{r}})/(0.5)^{\text{nr}^2+\text{r}}] [12], with nr referring to non-recombinants and r to recombinants assuming one allele is linked with the trait on the same chromosome, and with nr' and r' referring to the non-recombinants and recombinants, respectively, with the assumption that the alternative allele is initially on the same chromosome as the allele for the trait.$

Generation of a new Alutetraplex

Two Alutetraplexes had previously been generated [6,7] and reproduced in this study with the same protocols, but utilizing the safer GelRed stain. This work entails the development of a third tetraplex or AluTetraplex 3. Intermediate-frequency Alu elements and primers sets were selected from those identified by Roy-Engel et al. [3] and Carroll et al. [4]. The selections were based on having alleles that could be distinguished in size on a 2% agarose gel. Primer sets that consistently worked well together were utilized and additional sets then added until identifying four markers that generate interpretable and reproducible results. Alutetraplex 3 was generated by PCR amplification using 1X Buffer E (FailSafe buffer containing dNTPs and MgCl₂; Epicentre Biotechnologies, Madison, WI), primer (sequences provided in Table 1) concentrations of 270 nM (109F), 270 nM (109R), 270 nM (65F), 270 nM (65R), 230 nM (201F), 230 nM (201R), 340 nM (241F), 340 nM (241R), 1u FailSafe enzyme (Epicentre Biotechnologies) and 1 µl of crude prep DNA in a 25 µl volume using the following cycling

Alu I.D.	Forward Primer	Reverse Primer	+Alu	-Alu	Chr. #
Yb8NBC65	5'-atctcatctccctgcctctg-3'	5'-gggaggtctggagatctgtg-3'	517	186	6
Yb8NBC109	5'-gtgcctggtactctagaaataaaactctct-3'	5'-agaatgaaactccggctcaaa-3'	561	251	22
Yb8NBC201	5'-ggagaaaatgtaaggtttctagcac-3'	5'accaatgcaactatctacactgaca-3'	476	145	7
Yb8NBC241	5'-ggactgtgtctaagggtgtcct-3'	5'-ggtaattgggagcagttgaga-3'	427	97	7

Table 1: Loci and primers utilized for the generation of Alutetraplex 3 from Carroll et al. [4]. Sizes (bp) are given for amplification products of the presence and absence alleles, as well as the chromosome that harbors the locus.

conditions: 94°C for 2 m, followed by 32 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 1 m 30s, with a final extension step at 72°C for 5 m. Amplified products were separated on a 2% agarose gel in 1XTAE buffer with 1X GelRed stain (Biotium, Madison, WI) incorporated in the gel. Gels were run at 135 volts until the bromophenol blue band was at the bottom or just barely run off the gel.

Results

Refinements of Alu-based tetraplex reactions

Reproducible results of the two previously designed Alutetraplex reactions [7] were obtained (Figure 2) using rapidly extracted DNA preparations. For a laboratory course the BuccalAmp Extraction Kit is advantageous in generating ready to use DNA in just several minutes. However, since it is based only on heating to lyse the epithelial cells and degrade compounds inhibitory to amplification, then as with the commonly used Chelex method [1], the preparations are crude in relation to the purity of DNA obtained using affinity-based kits such as QIAamp. Crude preparations have been associated with lower stability and decreased amplification potential with long-term storage and multiple freeze thaws [17]. Due to the impurities, accurate quantitation could not be assessed using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Waltham, MA) as $\mathrm{OD}_{_{260}}/\mathrm{OD}_{_{280}}$ ratios of eleven samples isolated by the most recent class ranged from 1.05 to 1.38 (\overline{X} =1.13 \pm 0.09). Therefore, reactions incorporated the same volume of DNA rather than the same concentration. The samples used in the two Alubased tetraplexes had been stored for several years with multiple freeze thaws and retained interpretable results (Figure 2). Highly purified DNA isolated from peripheral lymphocytes [18], as anticipated, worked well in this study (data not shown). The much less toxic GelRed DNA stain was incorporated into these gels as opposed to ethidium bromide, with all eight alleles discernible in each tetraplex.

Incorporation of Alu-based tetraplexes in paternity analysis

Although *Alu*-based tetraplexes provide a unique tool to demonstrate paternity [7], increasing the number of biallelic markers improves the possibility of excluding individuals as the father, as well as enhances the verification of the correct individual. Individual







6 (Figures 2A and 2B) represents the maternal sample. Since she is heterozygous for all four Alu loci in tetraplex 1 (Figure 2A), then the data are uninformative for a paternity analysis. Therefore, the use of a second tetraplex was necessary to be able to perform the analysis (Figure 2B). All the alleles in the children (samples 7 and 8) that were not transmitted by the mother (6), can be accounted for by the father (sample 5). To help explain the concept, we could ask whether the individual represented by sample 9 could be the father. Since individual 7 has the 60- allele, whereas the mother (6) and individual 9 do not, then individual 9 is ruled out as the father of this child. We could rule out number 9 as the father of individual 8 in this case only because we can distinguish between two slightly different-sized 225+ alleles [9]. Individual 8 has the smaller 225+ allele (more details regarding this variant will be discussed) whereas neither the mother nor individual 9 does. If just labeling Alu 225 as + and - variants then we could not have ruled out individual 9 as the father of this child. The generation of a third *Alu*tetraplex therefore would enhance the paternity analysis by providing additional potentially informative markers.

Alu-based tetraplex reactions and their use in genetic mapping studies

Heterozygous markers are required in order to perform genetic mapping studies (outlined in Figure 3). However, heterozygosity of the mother in this pedigree is problematic, as she does not exhibit the trait, and heterozygous offspring from heterozygous parents would result in indistinguishable determination of the inherited paternal allele, which is being assessed for linkage to the trait. In this study Alutetraplex 1 was not informative since the mother was heterozygous for all markers (Figure 2A). However, *Alut*etraplex 2 provided markers informative for genetic mapping (Figures 2B and 4). In family 1, three of the four markers in the father (who has the trait) were heterozygous and the mother was homozygous for all four markers, allowing three genetic loci to be assessed for linkage. If the two children, both with the trait, shared the same paternal allele then the result would lean toward linkage, whereas if they had different paternal alleles then independent assortment would be inferred (Figure 3).

Genetic mapping using the *Alu* 9 dimorphism resulted in the children both obtaining the presence allele from the father (Figures 2B and 4), and Lod scores based on linkage at 10%, 20%, 30% and 40% (theta values in the formula) would yield values of 0.215, 0.134, 0.065, and 0.017, respectively. These are all positive numbers directed toward

Page 4 of 7



Figure 4: Two families analyzed for linkage of dimorphic *Alu* markers to the ACHOO syndrome. Numbering of individuals corresponds to Figure 2. Only informative data are presented. Shaded individuals have the syndrome.



Figure 5: Alutetraplex 3 developed in this study. PCR products were separated on a 2% agarose gel stained with GelRed. Lanes: M refers to molecular weight markers (Quick-Load PCR Markers, New England BioLabs, Inc., Ipswich, MA), numbers refer to anonymous DNA samples, -indicates the negative control (all reagents but DNA) and L refers to Quick-Load 100 bp DNA Ladder, with sizes indicated for bands up to 600 bp. Alleles are indicated on the left side of the figure.

linkage but represent too small a sample size to generate a significant value. In contrast the children had obtained different alleles for the Alu 60 and Alu 225 markers from the father, thereby generating Lod scores of -0.444, -0.194, -0.076, -0.018 for the same theta values, respectively. These negative values shift the inference toward non-linkage for these chromosomal locations. Values are additive and therefore as data are accumulated using different families, the Lod scores can be added, and values greater or equal to +3 provide significance that the marker is linked to the gene for the trait, whereas a value equal or less than -2 provides significant support that the marker is not linked to the gene. When analyzing the second family with the same marker set, the father was heterozygous for only Alu 225, but the mother was as well, with the

children also being heterozygous, and therefore none of the markers were informative. Additional markers, such as those incorporated into a third *Alu*tetraplex, increases the probability of demonstrating the additive effect by potentially generating informative data from more than one family.

Development of Alutetraplex 3

In this investigation a third Alu-based tetraplex was generated (Figure 5), consisting of markers 109, 65, 201, and 241(Table 1). The data were reproducible and variants easily discernible. As a rule-ofthumb, uncertainty can be resolved by noting the size of the fragment and the intensity of the alternative allele. A case in which this guideline can be used is sample 3 (Figure 5), as there is a barely visible band at approximately the size of the 201+ allele, although clearly a smaller fragment than that found in lanes 1, 4, and 5, but not as small as the notably visible 241+ allele. Additionally, the relative intensity of the 201allele for sample 3 is similar to sample 2, which is clearly homozygous for the absence allele. In accordance with this rationale is the comparison of the 109 alleles between samples 2 and 3, with 109- being displayed as a much more intense band in the homozygous sample 3 relative to the heterozygous sample 2. As verification of the correct interpretation, the data were consistent with individual testing of the primer sets (data not shown). This result was also consistent with repeated attempts using the markers 109, 65, and 201, which consistently worked well together. Several fourth markers were attempted (data not shown) to finalize this tetraplex. Marker 241 was found to consistently work well using the Failsafe system, with modifications of primer concentrations as previously described for building Alutetraplexes [6].

Alu-Based tetraplex reactions for forensic analysis and ancestry

Since two Alu-based tetraplexes have been generated, then 6,561 possible genotypes are possible. Therefore, in a laboratory course of 12 students, using one of the samples as the unknown, the identification of the unknown individual would be inevitable. To date, with over 60 samples used, no two individuals have had the identical genotype for the eight markers (data not shown). With more widespread use of this system, such as at various college and university classrooms, profiles can be submitted anonymously (e.g. http://www.emudnaprofiledatabase. org/index.php), increasing the data available for forensic analysis and could strengthen the potential use of this system for assessing ancestry. Also, the generation of additional Alu-based tetraplexes would exponentially increase the possible number of genotypes, providing a valuable asset in further pursuing these objectives. This would increase the complexity of the possible outcomes, while maintaining the simplicity of the technique. With the addition of four dimorphic Alu variants, the number of possible genotypes increases another 34. Twelve total markers on the three gels yield 531,441 possible genotypes. Genotypes and allele frequencies have been assessed in various populations (Table 2), supporting these markers as IF variants, hence effective as tools in forensic analysis. Based on the known values, students can then assess the possibility of generating their genotype. Hardy-Weinberg equilibrium is assumed for each population as none of the generated data statistically deviated from expected values. As an example of the potential for three Alu-based tetraplexes, the probability of obtaining the genotype 51A ++, 182 +-, ACE --, TPA25 +-, 9 --, 60--, 225 --, 50 +-, 109 +-, 65 ++, 201 +-, 241 --, using the hypothetical world value (based on average values of the populations tested) would be $(0.43)^2 \times 2$ (0.49) $(0.51) \times (0.49)^2 \times 2$ (0.43) $(0.57) \times (0.66)^2 \times (0.51)^2 \times (0.31)^2 \times 2$ (0.44) $(0.56)\times 2$ (0.36) $(0.64)\times (0.43)^2 \times 2$ (0.48) $(0.52)\times (0.82)^2$ or 1.67×10^{-6} .

Page	5	of	7
------	---	----	---

Population	Alu 51	Alu 182	ACE	TPA25	Alu 9	Alu 60	Alu 225	Alu 50	Alu 109	Alu 65	Alu 201	Alu 241
African American	+ 0.47 - 0.53	+ 0.33 - 0.67	+ 0.49 - 0.51	+ 0.30 - 0.70	+ 0.53 - 0.47	+ 0.59 - 0.41	+ 0.73 - 0.27	+ 0.05 - 0.95	+ 0.29 - 0.71	+ 0.56 - 0.44	+ 0.59 - 0.41	+ 0.00 - 1.00
Greenland natives	+ 0.42 - 0.58	+ 0.77 - 0.23	+ 0.55 - 0.45	+ 0.33 - 0.67					+ 0.33 - 0.67	+ 0.42 - 0.58	+ 0.39 - 0.61	
Asian			+ 0.30** - 0.70	+ 0.59** - 0.41	+ 0.48* - 0.52	+ 0.53* - 0.47	+ 0.72 - 0.28	+ 0.89* - 0.11				+ 0.13 - 0.87
European Caucasian	+ 0.48 - 0.52	+ 0.44 - 0.56			+ 0.35 - 0.65	+ 0.53 - 0.47	+ 0.70 - 0.30	+ 0.47 - 0.53	+ 0.41 - 0.59	+ 0.45 - 0.55	+ 0.58 - 0.42	+ 0.23 - 0.77
Egyptian	+ 0.35 - 0.65	+ 0.43 - 0.57	+ 0.68# - 0.32	+ 0.51 - 0.49	+ 0.00 - 1.00	+ 0.29 - 0.71	+ 0.60 - 0.40	+ 0.35 - 0.65	+ 0.39 - 0.61	+ 0.28 - 0.72	+ 0.34 - 0.66	+ 0.27 - 0.73
Total Estimate ^s	+ 0.43 - 0.57	+ 0.49 - 0.51	+ 0.51 - 0.49	+ 0.43 - 0.57	+ 0.34 - 0.66	+ 0.49 - 0.51	+ 0.69 - 0.31	+ 0.44 - 0.56	+ 0.36 - 0.64	+ 0.43 - 0.57	+ 0.48 - 0.52	+ 0.16 - 0.84

Table 2: Population Variation of Alu Dimorphisms [3,4,21,24,25].

*Article does not distinguish between Asian or Alaskan native populations tested

**Alaskan native values,

*control population in rheumatoid heart disease study

*The total values represent a hypothetical assumption of each population contributing equally to the world population (the average of the values)

Alu-based marker	Fst value (Egyptian-African American)					
51	0.030					
182	0.023					
50	0.285					
60	0.174					
225	0.035					
9	0.711					
201	0.228					
241	0.306					
109	0.022					
65	0.155					
ACE	0.074					
TPA25	0.091					

Table 3: Estimates of genetic differentiation between Egyptian and AfricanAmerican populations based on population variation in *Alu* dimorphism alleles(Table 2).

In contrast, the student can ascertain the probability of this genotype in his/her ethnic group. If African American ancestry was expected, the value would be 1.71×10^{-7} ; whereas if Egyptian, the expected value would be 1.07×10^{-6} . Albeit small data sets were evaluated, there is roughly an order of magnitude less of a probability that the individual is of African descent than Egyptian. Additionally, F_{st} calculations, measuring unbiased genetic distances between populations, depict three markers (*Alu* 50, *Alu* 241, *Alu* 9) associated with very high differentiation (values 0.25 or greater) between the African American and Egyptian populations (Table 3). The *Alu* 9 marker is exceptionally strong in distinguishing individuals of these populations. Therefore, *Alu*-based markers warrant consideration as applicable markers in depicting ancestry and ethnicity of an unknown sample.

Discussion

The *Alu*-based tetraplex methodology, incorporating the simultaneous analysis of four IF dimorphic *Alu* markers, provides a simple tool to generate complex DNA profiles. With basic laboratory equipment and the use of GelRed DNA stain, the system offers accessibility for a wide range of users, particularly in a genetics- or genomics-based laboratory course. Students can use a rapid extraction kit to utilize their own DNA for profiling, which has been shown to increase their interest in the material [19]. In this investigation a third *Alu*tetraplex was developed to increase the usefulness of this profiling tool while maintaining the simplicity of the system. With a third *Alu*tetraplex, this DNA profiling system is enhanced by increasing the possible number of genotypes to 531,441, along with minimal

additional procedural requirements that only include the obtaining of four additional primer sets and analyzing one additional PCR set. By utilizing three tetraplexes instead of 12 individual reactions for each sample, then time, reagents, and consumables are saved. Additionally, since the analyses are performed on mini agarose gels, then cost and time are further minimized. These are important considerations for a class-based laboratory. The generated profiles were demonstrated to be applicable to forensic, paternity, population and genetic mapping studies.

Ancestry or personalized genetic histories (PGHs) can be assessed by polymorphic DNA markers that substantially differ in allele frequencies across populations. These are referred to as ancestral informative markers (AIMs) [20]. A portion of Alu dimorphisms may be classified as AIMs based on distinct variation supported by weighted allele frequencies and Fst values between ethnic groups or populations of different geographic origins [21-24]. With a limited data set [3,4,21,24,25], very high genetic differentiation between African American and Egyptian populations was supported for three of the 12 Alu markers used in the tetraplexes. Additionally, when assessing a randomly selected genotype, the chance of the individual having an African American background was nearly an order of magnitude less than Egyptian. Combined, these findings illustrate the potential of this system to distinguish a sample among ethnic groups. The use of 100 Alu dimorphisms to infer geographic origins as either from Europe, Asia, Africa, or India, correctly assigned 14 of 18 tested unknown samples to one of these four major global populations and properly determined the other four as having mixed ancestry [26]. As anticipated, ancestry is not limited to one of four global populations. Additionally, Cordaux et al. [23] identified two dimorphic Alus in which the presence allele was restricted to African populations. The development of a highly comprehensive map of Alu and other mobile element insertion polymorphisms [27] provides continued building of population data advancing the power of these markers in assessing ancestry. Therefore, continued analysis of Alu dimorphisms and advancements of techniques to study these variants are anticipated to provide important contributions to the field.

The use of approximately 30 *Alu* dimorphisms [15,16] have been demonstrated to have resolving power approaching that of the Combined DNA Index System (CODIS) used by the FBI, involving 13 short tandem repeat loci (STRs) consisting of 8-22 alleles offering an approximate 10^{-15} probability of an exact match [28,29]. Advantages of of the *Alu* based system include the stability of *Alu* integrations and the lack of separate integrations into the exact same location making these

homoplasy-free markers, whereas STRs have the potential to change in number of repeats [30]. With additional Alu dimorphisms being identified with techniques such as anchored PCR [23], subtractive hybridization [31], and DNA sequencing of genomes [32], then Alubased systems have great potential to advance as a formidable tool in forensic analysis. One approach incorporates analyzing the Alu dimorphisms individually with PCR and gel electrophoresis [16]. This is a simple technique that provides clear results without specialized equipment and doable for standard molecular biology laboratories. An alternative novel method of Alu dimorphism analysis is the use of a multiplex PCR system using fluorescent dyes and is based on the target site duplication, in which size differences of 3-6 nucleotides analyzed by capillary gel electrophoresis distinguishes the alleles [15]. In contrast to these techniques, albeit thus far only 12 Alu variants are currently being analyzed, the tetraplex system provides a mechanism to analyze several variants at once, use less reagents than individual analysis of variants and less time, and does not require the use of specialized equipment or need for fluorescent dyes for labeling. The time and cost are important considerations for basic laboratories particularly in an educational setting. The development of a third Alu-based tetraplex escalated the profile possibilities and enhances this tool for use in paternity, genetic mapping, population, and forensic studies in addition to the potential use in assessing ancestry and/or geographic origins. Future development of a fourth profile of IF Alu markers would rocket the number of possible genotypes to 43,046,721 possible genotypes, gradually progressing to a range of potential commercial use.

A further advancement of the Alu-based system arose with the finding of additional Alu presence alleles that post-date the integration event [9]. Based on a small, but noticeable size variant of the Alu 225 marker as seen in Alutetraplex 2 (Figure 2B), which demonstrated a Mendelian pattern of inheritance for a three-allele system, then the presence allele was previously investigated further [9]. Interestingly, three distinct presence alleles were identified in addition to the absence form, and the evolutionary order of alleles could be assessed. A PCR with restriction enzyme digest protocol was developed to distinguish the four alleles. The system yielded additional findings not observed by presence/absence scoring. One presence allele was absent in the analyzed European Caucasians and South Americans, whereas a different presence allele was absent in all tested Asian individuals of various countries indicative of distinct geographic variation. The African American and various African populations exhibited all four alleles, consistent with the African origins hypothesis. This find suggested that there may be more informative allelic variation that post-dates Alu integration events, and this hypothesis is currently under investigation.

The application to genetic mapping was also enhanced by the development of the third *Alu*tetraplex by the provision of additional biallelic markers. Since not all biallelic markers generate informative data, as shown in this study, this advancement increases the possibility of identifying informative markers for students to determine Lod scores in order to assess linkage of the marker to the trait. In this case, mapping of the gene for the ACHOO syndrome was chosen since roughly 30% of the population have this non-health risk trait [14], and therefore identifying families to participate would be relatively easy. Since the determination of Lod scores is additive, then a proposed database can be generated for different laboratory courses to incorporate their data for the 12 different markers until significance for linkage or non-linkage is reached, hence students are participating and contributing real data in an ACHOO syndrome mapping consortium.

In conclusion, although large-scale projects and advanced technologies are being used to incorporate dimorphic *Alu* markers in forensic applications, the *Alu*tetraplex system with the newly developed third profile provides a tool that is relatively technologically simple and inexpensive, accessible for classroom use, and allows for highly informative forensics analyses along with an encouraging potential to assess ancestry. This simple method to generate complex profiles offers several advantages over the standard D1S80 VNTR and single *Alu* variant methods, particularly in an upper-level college course. Also presented was its use in genetic mapping to demonstrate positional cloning methodology. Lastly, this technique has the potential to advance further particularly with the growing number of identified dimorphic *Alu* elements.

Page 6 of 7

References

- Bloom MV, Freyer GA, Miklos DA (1996) Laboratory DNA Science. The Benjamin Cummings Publishing, Inc., Menlo Park, CA.
- Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3: 370-379.
- Roy-Engel AM, Carroll ML, Vogel E, Garber SV, Nguyen AH, et al. (2001) Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159: 279-290.
- Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, et al. (2001) Largescale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. J Mol Biol 311: 17-40.
- Thomas E, Herrera RJ (1998) Multiplex polymerase chain reaction of Alu polymorphic insertions. Electrophoresis 19: 2373-2379.
- Kass DH (2003) Generation of human DNA profiles by Alu-based multiplex polymerase chain reaction. Anal Biochem 321: 146-149.
- Kass DH (2007) Simple & rapid generation of complex DNA profiles for the undergraduate laboratory. American Bio Teach 69: 163-168.
- Kass DH, Laroe R (2007) Web-based analysis for student-generated complex genetic profiles. Biochem Mol Biol Educ 35: 404-409.
- Kass DH, Jamison N, Mayberry MM, Tecle E (2007) Identification of a unique Alu-based polymorphism and its use in human population studies. Gene 390: 146-152.
- Collins FS (1991) Of needles and haystacks: finding human disease genes by positional cloning. Clin Res 39: 615-623.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, et al. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409: 928-933.
- Primrose SB, Twyman R(2003) Principles of Genome Analysis and Genomics. (3rd edn), Wiley-Blackwell Publishing.
- Collie WR, Pagon RA, Hall JG, Shokeir MH (1978) ACHOO syndrome (autosomal dominant compelling helio-ophthalmic outburst syndrome). Birth Defects Orig Artic Ser 14: 361-363.
- 14. Peroutka SJ, Peroutka LA (1984) Autosomal dominant transmission of the "photic sneeze reflex". N Engl J Med 310: 599-600.
- Asari M, Omura T, Oka K, Maseda C, Tasaki Y, et al. (2012) Multiplex PCRbased Alu insertion polymorphisms genotyping for identifying individuals of Japanese ethnicity. Genomics 99: 227-232.
- Mamedov IZ, Shagina IA, Kurnikova MA, Novozhilov SN, Shagin DA, et al. (2010) A new set of markers for human identification based on 32 polymorphic Alu insertions. Eur J Hum Genet 18: 808-814.
- Greenspoon SA, Scarpetta MA, Drayton ML, Turek SA (1998) QIAamp spin columns as a method of DNA isolation for forensic casework. J Forensic Sci 43: 1024-1030.
- Kass DH, Batzer MA, Deininger PL (1996) Characterization and population diversity of interspersed repeat sequence variants (IRS-morphs). Genome 39: 688-696.
- 19. Daley LA, Wagner JK, Himmel TL, McPartland KA, Katsanis SH, et al. (2013)

Personal DNA testing in college classrooms: perspectives of students and professors. Genet Test Mol Biomarkers 17: 446-452.

- 20. Shriver MD, Kittles RA (2004) Genetic ancestry and the search for personalized genetic histories. Nat Rev Genet 5: 611-618.
- Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P, Papiha SS, Hammer M, et al. (2002) Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. Electrophoresis 23: 3346-3356.
- Romualdi C, Balding D, Nasidze IS, Risch G, Robichaux M, et al. (2002) Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. Genome Res 12: 602-612.
- Cordaux R, Srikanta D, Lee J, Stoneking M, Batzer MA (2007) In search of polymorphic Alu insertions with restricted geographic distributions. Genomics 90: 154-158.
- 24. Terreros MC, Alfonso-Sánchez MA, Novick GE, Luis JR, Lacau H, et al. (2009) Insights on human evolution: an analysis of Alu insertion polymorphisms. J Hum Genet 54: 603-611.
- 25. Morsy MM, Abdelaziz NA, Boghdady AM, Ahmed H, Abu Elfadl EM, et al. (2011) Angiotensin converting enzyme DD genotype is associated with development of rheumatic heart disease in Egyptian children. Rheumatol Int 31: 17-21.

 Ray DA, Walker JA, Hall A, Llewellyn B, Ballantyne J, et al. (2005) Inference of human geographic origins using Alu insertion polymorphisms. Forensic Sci Int 153: 117-124.

Page 7 of 7

- Stewart C, Kural D, Strömberg MP, Walker JA, Konkel MK, et al. (2011) A comprehensive map of mobile element insertion polymorphisms in humans. PLoS Genet 7: e1002236.
- Syn CK, Chuah SY, Ang HC, Lim SE, Tan-Siew WF, et al. (2005) Genetic data for the 13 CODIS STR loci in Singapore Chinese. Forensic SciInt 152: 285-288.
- 29. Ge J, Eisenberg A, Budowle B (2012) Developing criteria and data to determine best options for expanding the core CODIS loci. Investig Genet 3: 1.
- Laouina A, Nadifi S, Boulouiz R, El Arji M, Talbi J, et al. (2013) Mutation rate at 17 Y-STR loci in "Father/Son" pairs from moroccan population. Leg Med (Tokyo) 15: 269-271.
- Mamedov IZ, Arzumanyan ES, Amosova AL, Lebedev YB, Sverdlov ED (2005) Whole-genome experimental identification of insertion/deletion polymorphisms of interspersed repeats by a new general approach. Nucleic Acids Res 33: e16.
- 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, et al. (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073.