Inference of human geographic origins using \textit{Alu} insertion polymorphisms

David A. Ray\textsuperscript{a,1}, Jerilyn A. Walker\textsuperscript{a,1}, Ashley Hall\textsuperscript{b}, Barbara Llewellyn\textsuperscript{c}, Jack Ballantyne\textsuperscript{b}, Allen T. Christian\textsuperscript{d}, Kenneth Turteltaub\textsuperscript{d}, Mark A. Batzer\textsuperscript{a,*}

\textsuperscript{a}Department of Biological Sciences, Biological Computation and Visualization Center, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA
\textsuperscript{b}National Center for Forensic Science, University of Central Florida Post Office Box, 162367, Orlando, FL 32816-2367, USA
\textsuperscript{c}Illinois State Police, Forensic Science Center at Chicago, 1941 W. Roosevelt Road, Chicago, IL 60608, USA
\textsuperscript{d}Biology and Biotechnology Research Program, Lawrence Livermore National Laboratory, PO Box 808, L-452 Livermore, CA 94551, USA

Received 18 May 2004; received in revised form 26 October 2004; accepted 28 October 2004
Available online 7 April 2005

Abstract

The inference of an individual’s geographic ancestry or origin can be critical in narrowing the field of potential suspects in a criminal investigation. Most current technologies rely on single nucleotide polymorphism (SNP) genotypes to accomplish this task. However, SNPs can introduce homoplasy into an analysis since they can be identical-by-state. We introduce the use of insertion polymorphisms based on short interspersed elements (SINEs) as a potential alternative to SNPs. SINE polymorphisms are identical-by-descent, essentially homoplasy-free, and inexpensive to genotype using a variety of approaches. Herein, we present results of a blind study using 100 \textit{Alu} insertion polymorphisms to infer the geographic ancestry of 18 unknown individuals from a variety of geographic locations. Using a Structure analysis of the \textit{Alu} insertion polymorphism-based genotypes, we were able to correctly infer the geographic affiliation of all 18 unknown human individuals with high levels of confidence. This technique to infer the geographic affiliation of unknown human DNA samples will be a useful tool in forensic genomics.

Keywords: Forensic genomics; \textit{Alu}; Geographic affiliation; PCR

1. Introduction

Forensic DNA specimens are routinely matched to alleged criminal suspects in modern law enforcement. Frequently however, tools that narrow the potential pool of suspects are essential precursors to a positive identification in investigative forensics. The inferred ancestral origin of a DNA specimen is one type of evidence that can aid a criminal investigation. Human genetic variation and geographic population affiliation have been studied using many genetic systems, including mitochondrial [1–3], Y-chromosome [1,2], microsatellite [4,5], short tandem repeats (STR) [2,6–8], mobile elements [4,9–14], and single nucleotide polymorphisms (SNPs) [15–18].

Recently, Frudakis et al. [19] developed a SNP-based system for inference of ancestry for application to forensic casework. The initial system consisted of 56 SNP loci targeted from pigmentation and xenobiotic metabolism genes with ancestral diversity designed to identify individuals of European, African, and Asian descent [19]. Subsequently, Frudakis and DNAPrint\textsuperscript{TM} Genomics, Inc. (Sarasota, FL)
have introduced commercial applications of various SNP-based systems as a forensic service to law enforcement agencies. Notably, DNAWITNESS™ 2.0 was instrumental for inferring the geographic origin of the Louisiana serial killer in 2003 (www.dnaprint.com).

Although emerging SNP-based technologies have recently proven quite useful in law enforcement and will undoubtedly remain so in the future, SNPs have some limitations due to the fact that they represent single base pair differences. Like most other genetic polymorphisms, SNPs can be merely identical-by-state; that is, they may have arisen as a result of an independent parallel forward or backward mutation resulting in genotype misclassification (homoplas). Recent improvements in SNP-based approaches suggest that most of these problems can be overcome by carefully selecting the correct loci and the correct number of SNPs to use.

In this study, we introduce the use of insertion polymorphisms based on short interspersed elements (SINEs) as an alternative or a complement to existing systems. Mobile element insertion polymorphisms are essentially homoplas-free characters, identical by descent [20–23], and easy to genotype in a variety of formats [4,24–27]. The ancestral state of a human mobile element insertion polymorphism is known to be the absence of the element at a particular genomic location [10]. Alu elements are approximately 300 nucleotides in length and represent the most abundant class of short interspersed mobile elements in the human genome with more than one million copies [21]. Most of these elements are “fixed”, meaning that all individuals are homozygous for the insertion at a particular locus. However, members of several young Alu subfamilies such as Ya5, Ya8, Yb8, Yb9, Yc1, Yc2 and others, are polymorphic for insertion presence/absence [22,28–30] and different numbers of such markers have been shown to provide robust measurements of the relationships among various human populations [10,12,13,31]. These features make mobile element insertion polymorphisms virtual genomic fossils of ancestral lineage and, thus, a valuable tool for determining human geographic origins.

Here, we report the application of 100 Alu insertion polymorphisms as a forensic tool to ascertain the inferred geographic ancestry of unknown human DNA samples. In this blind study, we examined DNA specimens from 18 geographically diverse humans. For each sample, we used this blind study, we examined DNA specimens from 18 geographic ancestry of unknown human DNA samples. In this study, we introduce the use of insertion polymorphisms based on short interspersed elements (SINEs) as an alternative or a complement to existing systems. Mobile element insertion polymorphisms are essentially homoplas-free characters, identical by descent [20–23], and easy to genotype in a variety of formats [4,24–27]. The ancestral state of a human mobile element insertion polymorphism is known to be the absence of the element at a particular genomic location [10]. Alu elements are approximately 300 nucleotides in length and represent the most abundant class of short interspersed mobile elements in the human genome with more than one million copies [21]. Most of these elements are “fixed”, meaning that all individuals are homozygous for the insertion at a particular locus. However, members of several young Alu subfamilies such as Ya5, Ya8, Yb8, Yb9, Yc1, Yc2 and others, are polymorphic for insertion presence/absence [22,28–30] and different numbers of such markers have been shown to provide robust measurements of the relationships among various human populations [10,12,13,31]. These features make mobile element insertion polymorphisms virtual genomic fossils of ancestral lineage and, thus, a valuable tool for determining human geographic origins.

Here, we report the application of 100 Alu insertion polymorphisms as a forensic tool to ascertain the inferred geographic ancestry of unknown human DNA samples. In this blind study, we examined DNA specimens from 18 geographically diverse humans. For each sample, we used multi-locus genotypes from Alu insertion polymorphisms to infer geographic affiliation from among four major world populations.

2. Materials and methods

2.1. DNA samples

Eighteen anonymous human DNA samples were obtained under informed consent for this experiment by the Illinois State Police Forensic Science Center at Chicago and the National Center for Forensic Science, University of Central Florida in Orlando. The DNA from each sample was extracted from bloodstain cards or buccal swabs by the source laboratories (Illinois State Police and National Center for Forensic Science) and shipped to Louisiana State University (LSU) for genetic analysis using 100 Alu insertion polymorphisms [4,12] and a mobile element-based sex typing assay [26]. Investigators from each source laboratory had access to the physical description and geographic ancestry of the anonymous subjects while the analysis team at LSU remained blind to this data until the conclusion of the study.

2.2. Alu elements and PCR amplification

One hundred Alu insertion polymorphisms were used in this study. A complete list of the Alu elements oligonucleotide primers and amplification conditions is available as an electronic appendix (Supplementary Table 1) to this manuscript. It is also available at our website (http://www.batzerlab.lsu.edu) and at http://www.genome.org as supplemental material for Watkins et al. [4,12]. PCR reactions for agarose gel-based detection were carried out in 25 μl using 10 ng of DNA template, 1X PCR buffer II (Applied Biosystems, Inc.), 0.2 mM dNTPs, 200 nM each oligonucleotide primer, optimized MgCl₂, and one unit Taq DNA polymerase. Each sample was subjected to an initial denaturation of 1 min at 95 °C followed by 32 amplification cycles of denaturation at 95 °C for 30 s, optimized annealing for 30 s, followed by extension at 72 °C for 30 s. Amplicons were size-separated on a 2% agarose gel containing 0.2 μg/ml ethidium bromide and visualized by UV illumination (Fig. 1). Human gender identification was performed using sex chromosome specific mobile elements as previously reported by Hedges et al. [26].

2.3. Data analysis and Structure 2.0 inference

Genotypic data were recorded for each allele as follows: an individual who was homozygous present for a given Alu locus was assigned the code 1, 1; homozygous absent, 0, 0; and heterozygous, 1, 0. A sample of the data is shown in Table 1 (the complete data set is available in Supplementary Table 2 and at our website, http://batzerlab.lsu.edu under publication #152). The geographic affiliation of the samples was inferred using Structure 2.0 [32–34]. This software package performs model-based clustering using genotypic data from unlinked markers to infer population structure. For each individual, Structure 2.0 estimates the proportion of ancestry from each of K clusters. We used a burn-in of 15,000 iterations and a run of 20,000 replications. The sample size was 715 individuals of known geographic ancestry [12], plus eighteen individuals of unknown ancestry for a total of 733. Because previous analyses of the same known data indicated the presence of four distinct populations [4], the expected number of populations (K) was set at four (European, African, Asian, or Indian). Three replicate
runs were performed on the dataset, each requiring about 20 min using a desktop computer with a 3 GHz processor.

3. Results

In our analysis of the eighteen anonymous DNA samples, the amplification efficiency at each of the 100 Alu loci was 100%. Population assignment probabilities obtained from Structure 2.0 using the genotype data are outlined in Table 2. Of the 18 unknown samples, 14 were assigned to one population with a probability greater than 80% (N = 12 were identified as European, N = 1 was identified as African/African-American, and N = 1 was identified as Asian). The remaining four samples were classified as being of mixed ancestry (N = 3 admixture of European and African descent; and N = 1 admixture of Indian and Asian descent). Information revealed by the source laboratories following the study listed DNA samples #1–4, #7, #8 as European, and #5–6, #16 as African American. DNA sample #9 was listed as Jamaican, #10 of Greek ancestry, #11 as from Finland, #12 from England, #13 from Scotland, #14 from Italy, #15 from Venezuela, #17 from India, and #18 as Chinese. Our results for samples #10–14 suggested that

Table 1

Genotype data for 18 unknown DNA samples for nine of the 100 Alu loci used

<table>
<thead>
<tr>
<th>Locus #</th>
<th>1</th>
<th>1</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>3</th>
<th>4</th>
<th>4</th>
<th>5</th>
<th>5</th>
<th>6</th>
<th>6</th>
<th>7</th>
<th>7</th>
<th>8</th>
<th>8</th>
<th>9</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locus name</td>
<td>ACE</td>
<td>ACE</td>
<td>APO</td>
<td>APO</td>
<td>B65</td>
<td>B65</td>
<td>COL3A1</td>
<td>COL3A1</td>
<td>HS2.43</td>
<td>HS2.43</td>
<td>HS4.32</td>
<td>HS4.32</td>
<td>HS4.65</td>
<td>HS4.65</td>
<td>HS4.75</td>
<td>HS4.75</td>
<td>PV92</td>
<td>PV92</td>
</tr>
<tr>
<td>Subject 1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 7</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 9</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 15</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 16</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 17</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Subject 18</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

For each locus, there are two entries indicating the genotype of the sample. “1” indicates the presence of the Alu element at that allele and “0” indicates the absence of the element.
these were European in origin with a 92–99% probability. Sample #18 was identified as being of Asian descent with a 94% probability. Sample #15 tested as an admixture of 86% European/10% Asian, which is consistent with a Venezuelan origin.

The four samples classified as having mixed ancestry (<80% identity with one of the primary populations) were subjected to secondary analyses to obtain detailed admixture information. Based on Structure’s estimate of the most likely population(s) of origin, samples were assigned to each of the two potential source populations and admixture estimates were calculated for three parental generations. When samples #6 and #16 were assigned to Africa, the admixture analyses showed weak agreement that both were exclusively African (30% and 27%, respectively) with a 16–23% likelihood that at least one parent or grandparent was of European ancestry. Conversely, when #6 and #16 were assigned to the European population, there were strong indications of genetic contributions from an African parent for each subject, 99% and 76%, respectively. Both individuals were confirmed as African-American by the source laboratories. When sample #9 (Jamaican) was assigned to the European population, admixture analyses indicated a 1% likelihood that this was true, and a 47% probability that at least one grandparent or great-grandparent was of African descent. Conversely, when #9 was assigned to the African population, admixture analyses indicated a 6% likelihood that this was true, and an 87% probability that at least one parent was of European ancestry. Subject #17 (identified as from India by the source laboratory) showed the most admixture of the eighteen unknowns tested with strong affinity for both Indian (95%) and Asian (85%) populations, as well as a 24% probability that at least one great-grandparent was of European ancestry.

Variation in probability of assignment between the three original runs ranged from 0.1% to 7.9% (data not shown), with most (15/18) samples having a standard deviation of less than 0.012. The inferred geographic affiliation was consistent for all samples across the three runs. The standard deviation of population probability assignments among runs (average S.D. = 0.10) is shown for each sample in Table 2.

### 4. Discussion

The results of our study demonstrate the utility of this approach as a forensic tool. Determining the ancestry of an unknown human DNA sample could aid a criminal investigation by narrowing the pool of potential suspects. The Markov Chain Monte Carlo methodology used by the Structure 2.0 software package provides a powerful analysis to group all individuals into the selected number of populations and then determine the probability that each individual belongs to any given group [34]. In addition, the software has the ability to detect admixture between populations in individual genotypes going back several parental generations. We were successful in determining the geographic ancestry of the 18 unknown human DNA samples.
the probabilities of assignment were well over 80% and the detection of admixture in individuals of mixed ancestry was easily identified. Only one sample, #17 (see Table 2), gave results that might be considered ambiguous. However, given the complicated makeup of the Indian population [1], this result is not unexpected. Indeed, of the four populations in the current database of Alu insertion polymorphism variation, India is by far the most heterogeneous with many individuals clustering with either Europe or East Asia [4]. The results of our analyses were also consistent between runs suggesting that, in practice within investigative forensic laboratories, single runs of the analysis would be sufficient.

The 100 Alu insertion polymorphisms used in this study were largely mined from existing human genome databases [12]. However, since the human dispersal from Africa, Alu elements have continued to expand in the human genome [22,35]. For example, the more recent the insertion, the more likely it is to occur at high frequency in the geographic region of origin and exhibit very low allele frequencies elsewhere, thus, being indicative of its specific source population. The incorporation of additional population-indicative mobile element insertion polymorphisms to the existing panel of markers will eventually allow for subgroup (sub-continental) affiliation tests. We are in the process of implementing a cascade-like strategy to our method, which will consist of a series of tiered analyses for determination of “primary” geographic affiliation (Africa, Europe, Asia, or India), then for “secondary” or subgroup affiliation within each of these broad continental groups. Thus, once the initial Structure 2.0 analysis narrows the sample origin to a continental affiliation, subsequent analyses, using only insertion loci that are useful within one of these continental populations, have the potential to further isolate the unknown sample to sub-continental and regional origin. We are currently identifying additional mobile element insertion polymorphisms using PCR-based displays [36–38] and data mining [29] to identify sub-continental patterns of variation.

Previously, one limitation to this type of multiple locus approach has been that forensic DNA samples are often only available in trace quantities. The analysis of 100 separate PCR amplicons requires significantly more than trace amounts. Recent advancements in whole genome amplification (WGA) technologies such as RepliPHITM (EPICENTRE, Madison, WI) and GenomiPhiTM (Amersham Biosciences, Newark, NJ) have virtually eliminated this obstacle. Genomic DNA from residual cells left by incidental contact can be subjected to WGA and produce amplification patterns from the WGA templates which are completely consistent with the patterns observed using the original genomic DNA [39].

In an effort to confirm this for the 100 Alu insertion polymorphisms, we recently compared amplification patterns using original genomic DNA and WGA DNA. An aliquot of the original DNA was sent from LSU to LLNL where it was WGA using the method of Sorensen et al. [39] and then returned to LSU for comparative analyses. The genotypes were 97% (473 out of 489) consistent between the original DNA and the WGA DNA. Each of the 16 (of 489) disagreements (3%) represented a single allele aberration (i.e. between heterozygous and homozygous). The ability to determine the inferred ancestry of each individual was unaffected and was 100% consistent between the original and WGA DNA. The complete genotype results of this WGA experiment are presented in Supplementary Table 4 and on our website, http://batzerlab.lsu.edu under publication #152.

There are several advantages to the use of Alu insertion polymorphisms for the inference of human geographic origins. First, it can be a “low-tech” approach using standard PCR thermal cyclers and simple agarose gel electrophoresis commonly available in most laboratories. Second, Alu insertions are about 300 nucleotides long, identical by descent, and thus, quite stable compared to single nucleotide differences subject to forward or backward mutations. Furthermore, as more recent and more population-indicative Alu insertions are discovered and integrated into the analyses, the number of elements required to meet the needs of the investigator will decrease.

There are also disadvantages to this method. In most routine criminal investigations, inference of ancestry may be defined simply as Caucasian, African-American, or Asian, making our 100 Alu approach seem excessive and unnecessary. However, as law enforcement becomes increasingly global, the powerful statistical capability of our Alu-based approach using Structure will likely prove useful. Another disadvantage to our current method is that monoplex PCR reactions and manual electrophoretic systems can make the collection of genotype data prohibitively time consuming for a typical forensic laboratory. While the analysis of the Alu genotype data can be accomplished relatively quickly (<20 min on a 3 GHz processor), the development of multiplex compatible systems are essential for the transition of this approach to the forensic community. Although, multiplex PCR has been successful in testing three to four Alu element insertions simultaneously [27,40], at least 25 separate PCR reactions would still be required for data collection using these manual systems. Therefore, automated multiplexed genetic systems using high throughput analysis technology are currently under development in our laboratories. These involve fixing genomic DNA sequences representative of the “Alu present” sites and the pre-integration sites for the 100 Alu insertion polymorphisms such that DNA from an unknown individual can be screened using microplate or micro-array-based techniques.

Although there are pros and cons to every approach, the inference of an individual’s geographic ancestry is undoubtedly a useful bit of information when trying to narrow the pool of potential suspects during a criminal investigation. Here, we have presented results which demonstrate that analysis of 100 Alu insertion polymorphisms can be a powerful tool to accurately infer geographic origin. We believe that, together with currently accepted SNP-based
methods for identifying likely geographic ancestry (i.e. DNAWITNESS™ 2.0), this method can be a useful tool in forensic investigations. Furthermore, the eighteen anonymous human DNA samples used for this experiment were obtained directly from forensic science laboratories (Illinois State Police Forensic Science Center at Chicago and the National Center for Forensic Science, University of Central Florida in Orlando), illustrating the community’s interest in this approach. We anticipate that as this technique becomes increasingly refined, by adding new loci and implementing high throughput strategies for data collection, our Alu-based system for inference of human geographic origins will continue to evolve as a useful tool for forensic genomics.

Acknowledgments

This research was supported by an award from the Technical Support Working Group (M.A.B.) and by an award from the Central Intelligence Agency (K.T and A.T.C.). Work at Lawrence Livermore National Laboratory was conducted under the auspices of the U.S. Department of Energy under Contract No. W-7405-Eng-48.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at 10.1016/j.forsciint.2004.10.017.

References


between the LINEs: human genomic variation induced by LINE-1 retrotransposition, Genome Res. 10 (2000) 1496–1508.

