Alu Insertions and Genetic Diversity: A Preliminary Investigation by an Undergraduate Bioinformatics Class

Nancy L. Elwess*, Stephen L. Duprey, Lindesay A. Harney, Jessie E. Langman, Tara C. Marino, Carolina Martinez,Lauren L. McKeon, Chantel I.E. Moss, Sasha S. Myrie,Luke Ryan Taylor

Department of Biology, Plattsburgh State University, 101 Broad St., Plattsburgh, NY 12901

Email: nancy.elwess@plattsburgh.edu

*corresponding author

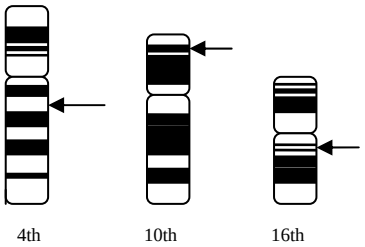
***Abstract*:** *Alu*-insertion polymorphisms were used by an undergraduate Bioinformatics class to study how these insertion sites could be the basis for an investigation in human population genetics. Based on the students' investigation, both allele and genotype *Alu* frequencies were determined for African-American and Japanese populations as well as a control. The three populations were tested for the presence of *Alu*-insertions on the 4th, 10th and 16th chromosomes.

***Keywords*:** bioinformatics, *Alu* insertions, SINES

Introduction

Alu elements, short interspersed elements (SINES) have been gathering within the human genome throughout the evolution of primates. *Alu* elements belong to a larger group of mobile elements which compose over 45% of our DNA (Batzer & Deininger, 2002). They are selfish pieces of DNA in that they don't encode for any proteins, they freely replicate and finally insert themselves into new chromosomal locations. These "jumping genes" are also known as transposable DNA [transposons] and were once studied in corn by Nobel Laureate, Barbara McClintock. *Alu* elements are specific to the primate genome and appeared roughly 65 million years ago (Carroll et al., 2001). With all of the replicating and raiding of chromosomes through insertions, *Alu* elements make up the largest of the SINES within humans; reaching over 1 million copies per genome and making up ~10% of the genome (Roy-Engel et al., 2001, Carroll et al., 2001). However this being said, it should be made clear that not all *Alu* insertions are the same. Once an *Alu* element becomes 'comfortable' in a new location, it starts to collect new mutations at the same pace as the surrounding DNA. Based on these new mutations, *Alu* elements are separated and organized into distinct lineages built on inheritance patterns. Since each *Alu* insertion is secure over evolution it is inherited by basic Mendelian genetics from parent to offspring. Consequently all individuals having an *Alu* insertion at a specific locus share a common ancestor from which they inherited the fragment. As a result, many of these *Alu* insertion sites are considered "landmarks" in the evolution of the human genome (Smit, 1996; Deininger & Batzer, 1999). Considering these factors, it was felt that the *Alu*-insertion polymorphisms would be an ideal topic of investigation in human population genetics for an undergraduate bioinformatics course. It was the intent of these bioinformatics' students to find differences in both the allele and genotype *Alu* frequencies by comparing two distinct ethnic populations (Japanese and African American) against a control. This was accomplished by using three different primers to detect specific *Alu* insertions on the 4th, 10th, and 16th chromosomes (Figure 1). Here, we present an analysis of our findings for these populations using *Alu* insertions.

Figure 1. *Alu* Insertion Sites on Chromosomes 4, 10 and 16.Arrows represent the approximate location of the *Alu* insertion on the respective chromosomes (National Center for Biotechnology Information). *Figure generated by Nancy L. Elwess.*



Methods and Materials

DNA Isolation Procedures (modified from the DNA Dolan Learning Center)

The students had the task of collecting over 60 DNA samples, this included ~20 samples from the control group; and each of the test groups (Japanese and African-American). The student investigators collected the samples from students on campus. Prior to the collection of samples, one liter of a 0.9% saline solution was made (9 grams of NaCl/1000 mL dH_2O). 10mL 0.9% saline solution was aliquoted into 50 mL polypropylene tubes.

To summarize the Dolan DNA Learning Center procedures, participants in this investigation were asked to swish 10 mL of 0.9% saline solution for approximately 30 seconds in their mouth, this was collected. In addition they were asked to sign a consent form. From each sample, one milliliter of the saliva-saline solution was placed into a 1.5 mL screw cap microcentrifuge tube and labeled with a number in order to identify the participant. The samples were concentrated for 1 minute at 12,000 rpm. A white pellet (containing cheek cells) resulted. The supernatant was removed and the pellet resuspended in 30 µl of saline solution.

To each resuspended sample 100 µl of 10% Chelex® was added. All the samples were placed in a boiling water bath for 10 minutes. The sample tubes were cooled on ice and spun for one minute at 12,000 rpm in a microcentrifuge. This step separated the DNA from the cellular debris. 30 µl of the top layer of supernatant from each sample tube was collected and transferred into a fresh 1.5 mL tube with the corresponding number, the resulting samples of DNA were used for the Polymerase Chain Reactions (PCR). The samples were stored on ice or placed in the freezer until they were needed for the PCR reactions.

DNA Amplification using Polymerase Chain Reaction

Reagents

Alu specific primers (Table 1) were ordered through Integrated DNA Technologies (IDT), each primer was diluted to a working concentration of 20 uM. The primers were designed to target regions upstream and downstream of a specific *Alu* insertion site. Each *Alu* fragment is approximately 300 base pairs in length. For example if there is no *Alu* insertion for Yb9NBC10 the size of the PCR product will be 197 base pairs, however if an *Alu* insertion is present the PCR product will be 524 base pairs for that primer.

Table 1. Primer sequences: Primer sequences for the three sets of *Alu* elements that were targeted for this study. Human diversity is classified as: High Frequency (HF) insertion polymorphism where the *Alu* element is present in all individuals tested except for one or two; Intermediate Frequency (IF) insertion polymorphism: the *Alu* element is present or absent in at least one population. We did not use any Low Frequency (LF) *Alu* insertion polymorphisms: these are *Alu* elements which are absent from all individuals tested except for one or two individuals.

Name	Primer Sequence	Chromosome Location	Human Diversity	Product Size (bp) With	Without
Yb9NBC10F Yb9NBC10R	5’ GTT TTC CTG GTG TGC CCT AAA TA-3’ 5’ TTT ACC TAA CTC ACA AGA CCC AAA G-3’	4	IF	524	197
Yc1NBC60F Yc1NBC60R	5’ GAAACCGCCAAGATTCTCACC -3’ 5’ TCTCCATCATGATTCCCAACTGA-3’	10	IF	522	205
PV92F PV92R	5’ GGA TCT CAG GGT GGG TGG CAA TGCT 5’GAA AGG CAA GCT ACC AGA AGC CCC A-3’	16	IF	731	416

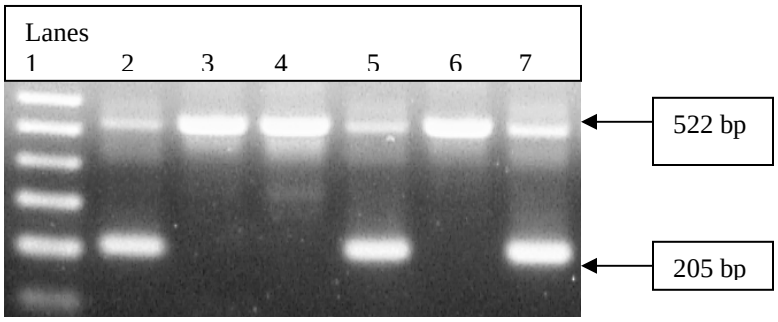
Procedures

Ready-to-Go PCR® tubes (GE Healthcare) were numbered with the corresponding sample numbers. To each tube, 17.5 ul of sterile dH_2O was added along with 2.5 ul of the desired forward and reverse primers for a specific *Alu* locus. Finally, 2.5 ul of each DNA sample was added to the corresponding numbered tube. The total final volume for each Ready-to-Go PCR® tube was 25 ul (17.5 ul of sterile dH_2O, 2.5 ul of each primer and 2.5 ul of DNA). Each sample tube was overlaid with 50 ul of mineral oil (since the thermal cycler did not have a heated lid) and added to the thermal cycler.
The thermal cycler was programmed for 30 cycles for the following cycle:
-Denaturing temperature and time: 94°C for 30 seconds
-Annealing temperature and time: 68°C for 30 seconds
-Extension temperature and time: 72°C for 30 seconds

DNA Gel Electrophoresis

20 µl of the amplified sample was retrieved from under the mineral oil for each sample and expelled into a new, labeled microcentrifuge tube containing 5 µl of loading dye. Each sample was loaded into a well on a 2% agarose gel (which contained 10 µl of 100 mg/mL ethidium bromide per 50 mL volume agarose). One lane was reserved for the 100 base pair ladder. Following electrophoresis, images of the gel were captured using a UV light box and a Kodiak gel documentation system, then interpreted (Figure 2)

Figure 2. A 2% agarose gel containing a 100 base pair standard (Lane 1) and six Japanese DNA samples (lanes 2-7) that were amplified with the Yc1NBC60 primers targeting chromosome 10. If the *Alu* insertion was present, a 522 base pair (bp) product was produced. If the *Alu* insertion was not present, then a 205 bp product was produced. Lanes 2, 5, and 7 contained the 10th chromosome's *Alu* target with the insertion and the 10th chromosome's *Alu* target without the insertion (+/-), hence the presence of two different sized bands. Lanes 3, 4 and 6, with only one band, designate these individuals as having the presence of two insertions (+/+) for the targeted *Alu* area.



Results

In this experiment the frequency of specific *Alu* insertions within different ethnicity groups were compared. Japanese and African-American groups were compared to the control group of random individuals for *Alu* insertions on the 4th, 10th and 16th chromosomes using the primers Yb9NBC10, YcNBC60 and PV-92 respectively. Table 2 shows the results for the genotype frequencies for each group and for each of the tested chromosome sites. If the insertion was present on both chromosomes, the individual was +/+, if the individual had one insertion on a chromosome and none on the homologous chromosome, then that person was +/-; finally, if no insertions were found on either chromosome then the person was -/-.

Table 2. Genotype frequencies; Genotype frequencies for the two test groups and the control group for the 4th, 10th, and 16th chromosomes. +/+ represents *Alu* insertions are present on both of the targeted chromosomes; +/- represents an *Alu* insertion is present on one of the targeted chromosomes; -/- represents that no *Alu* insertions are present for the targeted chromosomes.

	Control			**African-American**			**Japanese**		
Chromosome	+/+	+/-	-/-	+/+	+/-	-/-	+/+	+/-	-/-
4th	0%	18%	82%	0%	29%	71%	0%	0%	100%
10th	68%	14%	18%	33%	27%	40%	55%	36%	9%
16th	7%	13%	80%	0%	12.5%	87.5%	56%	13%	31%

The biggest differences that were seen for the genotype frequencies occurred on chromosomes 10 and 16. On chromosome 10, the majority of the Control group and Japanese groups had +/+ (68% and 56% respectively) compared to the African-American test groups with only 33% +/+. The Japanese test group did show differences from the Control and African-American groups for the 16th chromosome. The majority of the Japanese had +/+ (56%) while the Control and African-American groups had 7% and 0% respectively for +/+.

Figures 3-5 provide the results for the allele frequencies for the targeted 4th, 10th and 16th chromosome *Alu* insertion sites. The allele frequency was determined by comparing the number of copies for a specific allele to the total number of alleles present. For example if the results had 10 +/+ individuals, 5 +/- individuals and 5 -/- individuals, the allele frequency would have a total of 25 + (insertions) and 15 – (no insertions). This would result in a 62.5% + and 37.5% - allele frequency. There was not that big of a difference in allele frequency between the three groups for chromosome 4 (Figure 3). However, there were differences in allele frequencies on Chromosome 10 for the African-American group (Figure 4) and on Chromosome 16 for the Japanese test group (Figure 5).

Figure 3. Distribution of allele frequencies for *Alu* insertions on the 4th chromosome.

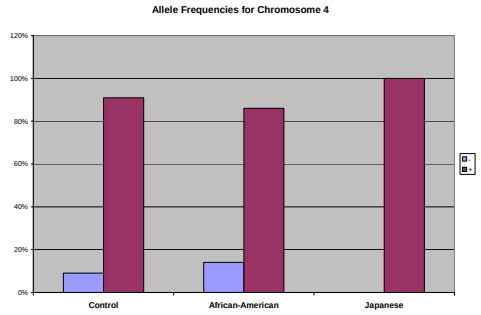
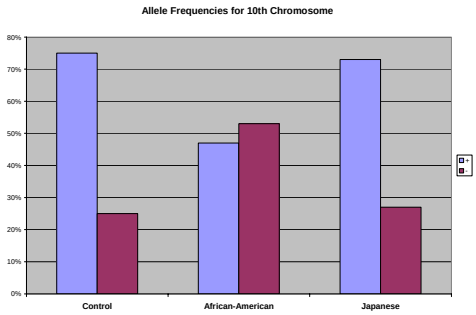


Figure 4. Distribution of allele frequencies for *Alu* insertions on the 10th chromosome.



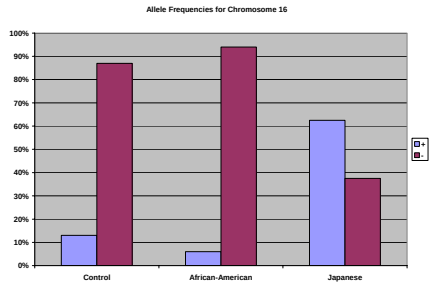
Discussion

Comparisons to data from past research and the literature can also be drawn from this experiment. For example, the Dolan DNA Learning Center (www.geneticorgins.org) provides a database of allele and genotype frequencies for the PV-92 *Alu* insertion (on the 16th chromosome) from over 40 populations around the world. According to the database the African-American allele frequency for this *Alu* insertion was 20%, whereas only 6% of our African-American test group had this insertion present. However, this discrepancy in the data may be due to the small population size for not only our study (21 samples) but also in the Dolan DNA Learning Center database (42 samples). More sampling using the African-American population, needs to be done to establish a reliable set of data.

The Dolan DNA Learning Center database (and other sources) did not have any data on the frequencies of the PV-92 *Alu* insertions in the Japanese population. So whatever data we could provide added to the knowledge in the field. However, the Japanese population was compared to other Asian

populations, which were known to have a high et al., 2001). There was a 90% + allele frequency for the Taiwanese population, 86% for the Chinese, 80% for the Filipino population as compared to the 62.5% + allele frequency in our Japanese test group.

Figure 5. Distribution of allele frequencies for *Alu* insertions on the 16th chromosome.



Further comparisons were made to published results of the Yb9NBC10 (Chromosome 4) and Yc1NBC60 (Chromosome 10) *Alu* elements (Roy-Engel et al., 2001). According to the published results the African-American test group had 37.5% +/+, 12.5% +/-, and 50% -/- for the Yb9NBC10 *Alu* element this frequency of the *Alu* insertion at this site (Comas was different from our results of 0% +/+, 29% +/-, and 71% -/-. It should be noted, however, that the Roy-Engel et al. article (2001) findings were based on only 8 samples compared to our 21 samples. When comparisons for the African-American populations were made concerning the Yc1NBC60 *Alu* element between the Roy-Engel paper and our study, there were once again differences. We had 33% +/+, 27% +/-, and 40% -/- for our African-American samples compared to their results of 33.33% +/+, 50% +/-, and 16.66% -/-. Here again the Roy-Engel paper had a smaller sample size then our study.

We could not make these direct comparisons for our Japanese results. The only published results that were close were for Asian/Alaska natives. This made our findings even more exciting due to no other published results for a Japanese test group.
Finally, in addition to our *Alu* findings, the bioinformatics students researched and presented findings from journal articles about genetic disorders/diseases that happen as a result of an *Alu* insertion within a gene. If time had allowed, we would have added more populations to sample and additional *Alu* locations to study.

References

BATZER, M.A., STONEKING, M., ALEGRIA-HARTMAN, M., BAZAN, H., KASS, D., SHAIKH, T., NOVICK, G., AND IOANNOU, P.A. 1994. African origin of human-specific polymorphic Alu insertions. *Proc. Natl Acad Sci* 91, 12288-12292.

BATZER, M.A., DEININGER, P.L., HELLMANN-BLUMBER, U., JURKA, J., LABUDA, D., RUBIN, C.M., SCHMID, C.W., ZIETKIEWIICZ, E., AND ZUCKERKANDL, E. 1996. Standardized nomenclature for Alu repeats. J. Mol. Evol. 42, 3-6.

BATZER, M.A. AND DEININGER, P.L. 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3, 370-379.

CARROLL, M.L., ROY-ENGEL, A.M., NGUYEN, S.V., SALEM, A-H., VOGEL, E., VINCENT, B., MYERS, J., AHMAD, Z., NGUYEN, L., SAMMARCO, M., WATKINS, W.S., HENKE, J., MAKALOWSKI, W., JORDE, L.B., DEININGER, P.L., AND BATZER, M., A. 2001. Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. Journal of Molecular Biology 311, 17-40.

COMAS, D., PLAZA, S., CALAFELL, F., SAJANTILA, A., AND BERTRANPETIT, J. 2001. Recent insertion of an *Alu* element within a polymorphic human-specific *Alu* insertion. Molecular Biology and Evolution, 18, 85-88.

DEININGER, P.L. AND BATZER, M.A. 1999. Alu repeats and Human Disease. Molecular Genetics and Metabolisim, 67, 183-193.

DOLAN DNA LEARNING CENTER, COLD SPRING HARBOR LABORATORY. Alu insertion polymorphism. Retrieved August 24, 2007, from http://www.geneticorigins.org

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION.2007. Human Genome. Retrieved August 26, 2007 from http://www.ncbi.nlm.nih.gov/

ROY, A.M., CARROLL, S.V., NGUYEN, S.V., AND SALEM, A-H.1999. Recently integrated

human Alu repeats: finding needles in the haystack.

ROY-ENGEL, A.M., CARROLL, M.L., VOGEL, E., GARBER, R. K., NGUYEN, S.V., SALEM, A-H., BATZER, M.A., AND DEININGER, P.L.2001. Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159, 279-290.

SMIT, A.F. 1996. The origin of interspersed repeats in the human genome. Curr. Opin. Genet. Dev. 6, 743-748.