Alu Dimorphism at the PV92 Locus of Chromosome 16 is in Equilibrium for University Student Population

James M. Brooks* and Mackenzie L. Thackston

Department of Biology, Charleston Southern University, 9200 University Boulevard, Charleston, SC 29406

Alu is a retrotransposable element, which refers to its ability to be copied and move from one region of DNA to another DNA region. At the PV92 locus of chromosome 16, *Alu* is a 300 bp dimorphic insert that can either be present or absent. It does not encode a protein product and has lost the ability to transpose. It is specific to humans, and differences in genotype and allele frequencies between human populations are important tools in understanding evolution. In this research, data was obtained and analyzed from 269 students at Charleston Southern University (CSU) belonging to four different races: Asian, Black, Hispanic/Latino, and White. Standard molecular biology procedures were used to isolate DNA from epithelial cheek cells, detect *Alu* inserts using polymerase chain reaction (PCR), and determine genotypes by gel electrophoresis. Statistical analyses were performed using Microsoft Excel, and chi square and Hardy-Weinberg equations were used to test for goodness of fit and equilibrium, respectively. The results were separated by genotypes: homozygous present, heterozygous, or homozygous absent. Homozygous absent was the most common genotype. Results were further separated into categories of gender and race. No significant genotype differences were found between male and female or between Black and White students. Nevertheless, there were significant differences between all other race combinations. Hardy-Weinberg calculations indicate that mutations, natural selection, nonrandom mating, genetic drift, and gene flow are negligible, and the overall student population at CSU is in equilibrium.

Introduction

Alu is an example of a polymorphism in the human genome. It is a retrotransposable element, which refers to its ability to be copied and move from one region of DNA to another DNA region. It is also referred to as a jumping gene, but does not encode a protein product and may exist only for its own replication.[1] There are many transposable elements that are specific to different organisms; however, *Alu* is specific to primates and the PV92 locus of chromosome 16 to humans. *Alu* is a member of the family of short, interspersed elements (SINEs). It is approximately 300 nucleotides in length, but with an estimated one million copies on a person's chromosomes, it makes up about 11% of the human genome.[2] At the PV92 locus of chromosome 16, *Alu* can be either present or absent, and has lost the ability to transpose. An individual may test homozygous present having the *Alu* on two chromosomes, heterozygous having the *Alu* on one chromosome, or homozygous absent lacking PV92 *Alu* on both chromosomes. *Alu* is believed to have inserted into the human genome during the last million years during dispersion of modern humans.[3] Therefore, differences in genotype and allele frequencies between human populations are an important tool in understanding human evolution.

Alu insertion polymorphisms are excellent markers for studying the genetic structure and relationships among human populations.[4-9] Several reviews focusing on *Alu* elements have been published.[10-13] Although most *Alu* inserts do not cause disease, some deleterious *Alu* activity is associated with a number of disorders in humans.[2, 14-17] The *Alu* insert at the PV92 locus of chromosome 16 along with other *Alu* genetic marker locations are widely used to estimate genetic diversity of human populations. In addition to other geographic locations, researchers have examined PV92 polymorphisms in populations from Africa[5], Russia[18], the Caucasus[19], England[20], Spain[21], Eurasian Regions[22], South Morocco[23], Argentina[24], Malaysia[25], Uruguay[26], Ivory Coast[27], Siberia[28], and Nigeria[29]. One preliminary study examined the PV92 locus of 60 university students from African-American and Japanese populations as well as a control.[30] Nevertheless, to the best of our knowledge no large-scale study has been published on the PV92 locus of chromosome 16 for a university student population or focused on Hardy-Weinberg equilibrium.

Hardy-Weinberg equilibrium is a state where allele and genotype frequencies in a population remain constant from generation to generation. In Hardy-Weinberg equilibrium populations are not evolving in terms of a specific gene. There are certain factors that can affect equilibrium including mutations, natural selection, nonrandom mating, genetic drift, and gene flow.[31] Small populations are more susceptible to changes and are typically not in equilibrium. Large populations are more likely to be in equilibrium, as the effects of mutations, mating, and genetic drift are negligible. Nevertheless, disequilibrium is still frequent in large populations because natural selection and gene flow are common.

This research was performed at Charleston Southern University (CSU), which is a liberal arts four-year private university in North Charleston, South Carolina. According to the school's website[32], as of 2021, there were approximately 3,350 students enrolled. Roughly 15% of the undergraduate students are age 25 and older. The international student population makes up about 2% of the student body and comes from 41 different countries. The full-time CSU undergraduate population is made up of 64% females and 36% males. CSU has a range of ethnicities represented on campus, which according to Data USA[33] include White (59.6%), Black (19.8%), Hispanic/Latino (4.1%), Asian (1.6%), and other or unknown (14.9%). The purpose of this research is to find the PV92 *Alu* genotypes of a racially diverse university student population and determine if it is in equilibrium. We obtained and analyzed data from 269 students at CSU belonging to four different races; Asian, Black, Hispanic/Latino, and White.

Methods

Sample collection

All samples came from undergraduate students enrolled in an introductory genetics lab course at CSU. DNA was collected at random from volunteers. Samples were not linked to specific individuals, but were grouped by gender and race, which was determined by phenotypic appearance. This research was approved by the Institutional Review Board at CSU.

DNA isolation

DNA was isolated following the methods established by Dolan[34] with a few modifications. Cells were collected by vigorously rinsing cheek pockets for 30 seconds with 10 mL of 0.9% saline solution. A 1.5 mL aliquot was then centrifuged at 5,000 rpm for 90 seconds to pellet the cells. The supernatant was removed, and the pellet was suspended in 30 µL of 0.9% saline solution. The sample was then placed into a 1.5 mL tube along with 100 µL of 10% Chelex resin (Carolina Biological Supply Company) and heated for 10 minutes at 95 °C in a heat block. Heat lysed the cells, and the Chelex resin removed metal contaminating ions. After heating, the sample was vigorously shaken for 5 seconds and centrifuged at 5,000 rpm for 90 seconds. DNA was collected by taking 30 µL of the clear supernatant and stored at -20 °C until used in the next step.

Polymerase chain reaction (PCR)

PCR was performed using PuReTaq Ready-To-Go™ PCR Beads with a primer loading dye mix from Carolina Biological Supply. Primer sequences were 16S-F: 5'- GGATCTCAGGGTGGGTGGCAATGCT-3' and 16S-R: 5'-GAAAGGCAAGCTACCAGAAGCCCCAA-3'. A 2.5 µL aliquot of isolated DNA was mixed with 22.5 µL primer loading dye mix and one Ready-To-Go™ Bead per sample. The thermocycler was programmed for an initial denaturation of 2 minutes and 94 °C, followed by 35 cycles of denaturation for 1 minute at 94 °C, hybridization for 1 minute at 68 °C, and extension for 2 minutes at 72 °C, with a final extension of 5 minutes at 72 °C. PCR products were stored at -20 °C until used in the next step.

Gel electrophoresis

PCR products were separated on a 1.5% agarose gel by electrophoresis along with a pBR322/*Bst*NI DNA ladder (Carolina Biological Supply Company) and stained with SYBR™ Safe DNA gel stain (Invitrogen™) using previously established protocols.[35] Gels were visualized and photographed using a Gel Doc™ XR system (Bio-Rad Laboratories). A 4 µL aliquot of a positive control was added to the far left lane and a 4 µL aliquot of negative control to the next lane of each gel. In separate wells 10 µL of PCR product were added per sample, as described by Dolan.[34]

Results

In this research, 269 individuals were tested for their PV92 *Alu* genotypes. Table 1 shows the data organized by race, gender, and genotype. It reveals the majority of students are White, there are more females than males tested, and homozygous absent is the most common genotype. Figure 1 is an example of a standard gel, illustrating each of the genotypes and controls. The *Alu* insert can be present (+) or absent (-) on each person's two chromosomes. If absent, a DNA band of about 380 bp is seen. If present, a larger DNA band around 680 bp is seen because the *Alu* insert is approximately 300 bp in length. Figure 2 shows the percent of individuals with each of the three genotypes to compare the numbers within and between races and genders. Comparing the genotypes in terms of percent allows for an easier evaluation of trends since each category has a different number of participants. Asian students have the highest percent of homozygous present, Hispanic/ Latino students have the highest percent of heterozygous, and White students have the highest percent of homozygous absent genotypes. Other apparent trends are the similarity in the distribution of genotypes between male and female White students and between Black and White female students. Goodness of fit between genders and races was determined using chi square tests. Analyses indicated no significant differences between males and females or between Black and White students, but there was a significant difference between other races (Table 2). Hardy-Weinberg calculations (Table 3) suggested the student population is close to equilibrium, and chi square analysis showed no significant difference between the CSU student population tested and equilibrium (Table 2).

Discussion and Conclusions

Figure 1 shows a standard gel, demonstrating each of the genotypes and controls. The SYBR™ Safe DNA gel stain used is more sensitive than traditional stains such as ethidium bromide or CarolinaBLU™. Therefore, faint bands of the incorrect size caused by nonspecific amplification were observed and ignored when interpreting the gels. It was also common to see insoluble material remaining in the wells and diffuse smaller sized bands towards the bottom of the gel caused by primer dimmers. However, only prominent bands of the correct sizes were used to determine PV92 *Alu* genotypes.

The data is difficult to compare in terms of numbers, as the majority of participants in the research were White females (Table 1). Nevertheless, the data collected is representative of the student population at CSU. After comparing the genotypes in terms of percent (Figure 2), chi square calculations were used to see if any relationships were a good fit between the races or genders. When comparing males

Table 1. Data organized by race, gender, and genotype.

Race	Gender	Genotype (-,-)	Genotype (+,-)	Genotype (+,+)	Totals
Asian	Male	2	0	1	3
	Female	3	2	2	7
	Genotype Total	5	2	3	10
Black	Male	3	6	1	10
	Female	16	12	1	29
	Genotype Total	19	18	2	39
Hispanic or Latino	Male	0	2	0	2
	Female	1	3	1	5
	Genotype Total	1	5	1	7
White	Male	34	18	4	56
	Female	88	57	12	157
	Genotype Total	122	75	16	213
		Total Participants:			**269**

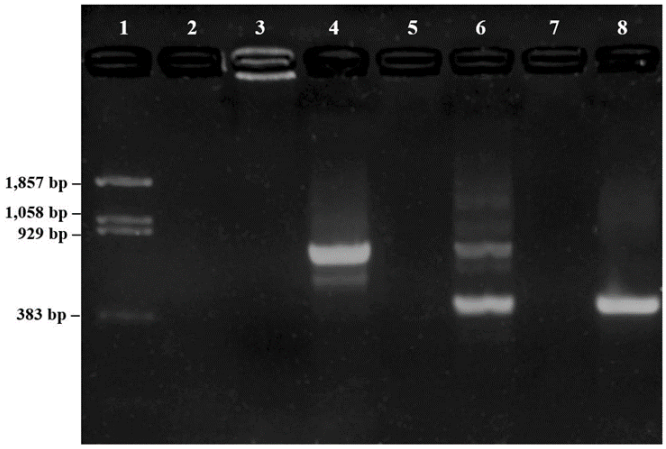

Figure 1. Example of a gel showing controls and three possible genotypes. Lane 1 shows a positive control (marker pBR322/ *Bst*NI), lane 3 shows a negative control (no DNA), lane 4 shows a (+, +) genotype, lane 6 shows a (+, -) genotype, and lane 8 shows a (-, -) genotype.

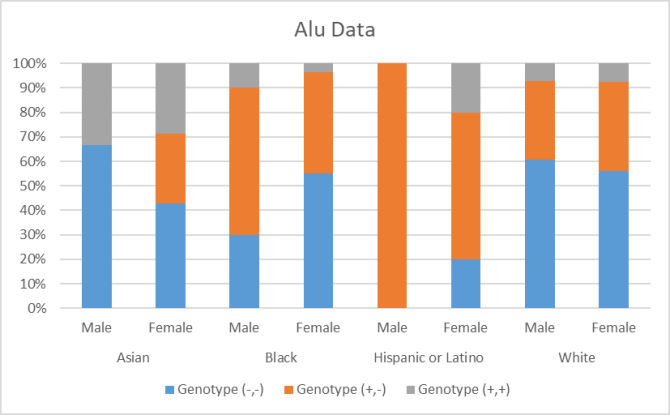

Figure 2. Percent of each genotype organized by race and gender.

Table 2. Goodness of fit for gender and race data.

Genotype Comparisons	Ratio	Chi Square	Significant Difference
Male vs Female	0.358	0.0817	No
Asian vs Black	0.256	20.88	Yes
Asian vs Hispanic or Latino	1.429	23.571	Yes
Asian vs White	0.047	61.917	Yes
Black vs Hispanic or Latino	5.571	34.41	Yes
Black vs White	0.183	1.88	No
Hispanic or Latino vs White	0.033	313.26	Yes
Observed vs Hardy-Weinberg	1.000	0.26	No

Table 3. Calculations for frequencies and equilibrium.

	Genotype Frequency	**Allele Frequency**	**Hardy-Weinberg Equilibrium**
(-,-)	54.60%	N.A.	53.60%
(+,-)	37.20%	N.A.	39.20%
(+,+)	8.20%	N.A.	7.20%
-	N.A.	73.20%	N.A.
+	N.A.	26.80%	N.A.

and females the chi square had a value of 0.0817 (Table 2). This reveals that even though the number of males to females is not even, the ratio of each genotype was close to equal. However, none of the races had a similar genotype distribution except for the Black and White races. This chi square value is 1.88 (Table 2). The other races in comparison to each other resulted in chi square values ranging from approximately 20-300 (Table 2), which shows they are significantly different. Nevertheless, this data should be interpreted with caution because of the small sample size of Asian and Hispanic/Latino participants.

The data presented here shows that the population of students at CSU campus is close to Hardy-Weinberg equilibrium (Table 3). The five main factors that can cause disequilibrium are mutations, natural selection, nonrandom mating, genetic drift, and gene flow. One or more of these factors could have thrown off the equilibrium on campus. There is no evidence of new mutations at the PV92 locus. Since the *Alu* sequence studied is phenotypically neutral, natural selection is probably not selecting for or against it, and participants mate randomly in relation to PV92 genotypes. Genetic drift is likely negligible, as the participants represent a large population. Therefore, gene flow is the most likely candidate to cause disequilibrium, as CSU has a racially diverse student population from multiple states and countries. Nevertheless, overall the genotype frequencies (Table 3) observed were in Hardy-Weinberg equilibrium, with a chi square value of only 0.26 (Table 2).

This research is significant as it reveals trends among races and genders. *Alu* inserts are useful for understanding human migration and evolution, as they have different allele frequencies among races. This research found that the highest PV92 *Alu* allele frequencies are in Asian and Hispanic/Latino populations (Figure 2), which should be verified by additional research using larger sample sizes, even though it is consistent with other studies.[5, 25, 34] It would also be interesting to confirm no new mutations occurred by sequencing PCR products from heterozygous genotypes. Although PV92 *Alu* has no known connection to disease, other *Alu* sequences can be tied to 0.4% of human genetic diseases.[36] Therefore, further research on *Alu* is important to understand certain disorders in humans and work toward cures and prevention.

In conclusion, this is likely the first large-scale study that has been published on the PV92 locus of chromosome 16 for a university student population. Among the 269 students tested at CSU, homozygous absent is the most common genotype. Asian students have the highest percent of homozygous present, Hispanic/Latino students the highest percent of heterozygous, and White students the highest percent of homozygous absent genotypes. Consistent with other studies[5, 25, 34], the highest PV92 *Alu* allele frequencies are in Asian and Hispanic/Latino populations. No significant genotype differences were found between male and female or between Black and White students. Nevertheless, there are significant differences between all other race combinations. Hardy-Weinberg calculations indicate that the overall student population at CSU is in equilibrium.

Acknowledgements

Charleston Southern University Biology Department generously provided funding. Undergraduate volunteers freely donated samples. Imelda T. Brooks and Alaina C. Urman kindly helped with the proofreading and editing of this manuscript.

Notes and References

**Corresponding author email: jbrooks@csuniv.edu*

1. Pray LA. 2008. Functions and utility of *Alu* jumping genes. *Nature Education* 1: 93.
2. Ade C, Roy-Engel AM, and Deininger PL. 2013. *Alu* elements: an intrinsic source of human genome instability. *Current Opinion in Virology*. 3: 639-645.
3. DNA Learning Center. 2013. Genetic origins: *Alu* insert polymorphism. Cold Spring Harbor Laboratory (New York). [cited 18 June 2022]. Available from: http://www.geneticorigins.org/pv92/aluframeset.htm
4. Batzer MA, Stoneking M, Alegria-Hartman M, Barzan H, Kass DH, Shaikh TH, Novick GE, Iannou PA, Scheer WD, Herrera RJ, and Deininger PL. 1994. African origin of human-specific polymorphic *Alu* insertions. *Proceedings of the National Academy of Sciences*. 91: 12288-12292.
5. Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, and Batzer MA. 1997. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research*. 7: 1061-1071.
6. Comas D, Calafell F, Benchemsi N, Helal A, Lefranc G, Stoneking M, Batzer MA, Bertranpetit J, and Sajantila A. 2000. *Alu* insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Straits. *Human Genetics*. 107: 312-319.
7. Terreros MC, Martinez L, and Herrera RJ. 2005. Polymorphic *Alu* insertions and genetic diversity among African populations. *Human Biology*. 77: 675-704.
8. Ennafaa H, Amor MB., Yacoubi-Loueslati B, Khodjet el-Khil H, Gonzalez-Perez E, Moral P, Maca-Mayer N, and Elgaaied A. 2006. *Alu* polymorphisms in Jerba Island population (Tunisia): comparative study in Arab and Berber groups. *Annals of Human Biology*. 33: 634-640.
9. Frigi S, Ennafaa H, Amor MB, Cherni L, and Ben Ammar-Elgaaiod A. 2011. Assessing human genetic diversity in Tunisian Berber populations by *Alu* insertion polymorphisms. *Annals of Human Biology*. 38: 53-58.
10. Rowold DJ and Herrera RJ. 2000. *Alu* elements and the human genome. *Genetica*. 108: 57-72.
11. Batzer MA and Deininger PL. 2002. *Alu* repeats and human genomic diversity. *Nature Reviews Genetics*. 3: 370-379.
12. Hasler J and Strub K. 2006. *Alu* elements as regulators of gene expression. *Nucleic Acids Research*. 34: 5491-5497.
13. Deininger P. 2011. *Alu* elements: know the SINEs. *Genome Biology*. 12: 236.
14. Ule J. 2013. *Alu* elements: at the crossroads between disease and evolution. Biochemical Society transactions. 41: 1532-1535.
15. Larsen PA, Hunnicutt KE, Larsen RJ, Yoder AD, and Saunders AM. 2018. Warning SINEs: *Alu* elements, evolution of the human brain, and the spectrum of neurological disease. *Chromosome Research*. 26: 93-111.
16. Li S, Zong L, Hou Y, Zhang W, Zhou L, Yang Q, Wang L, Jiang W, Li Q, Huang X, Ning Y, Wen Z, and Zhao C. 2019. Altered DNA methylation of the *Alu*Y subfamily in schizophrenia and bipolar disorder. *Epigenomics*. 11: 581-586.
17. Burns KH. 2020. Our conflict with transposable elements and its implications for human disease. *Annual Review of Pathology*. 15: 51-70.
18. Stepanov VA, Puzyrev VP, Spiridonova MG, and Khitrinskaia IY. 1999. Analysis of the *Alu* insertion polymorphism in urban and rural Russian populations of Siberia. *Russian Journal of Genetics*. 35: 1138-1143.
19. Nasidze I, Risch GM, Robichaux M, Sherry ST, Batzer MA, and Stoneking

M. 2001. *Alu* insertion polymorphisms and the genetic structure of human populations from the Caucasus. *European Journal of Human Genetics*. 9: 267-272.
20. Mastana S, Lee D, Singh PP, and Singh M. 2003. Molecular genetic variation in the East Midlands, England: analysis of VNTR, STR and *Alu* insertion/ deletion polymorphisms. *Annals of Human Biology*. 30: 538-550.
21. Garcia-Obregon S, Alfonso-Sanchez MA, Perez-Miranda AM, Vidales C, Arroyo D, and Pena JA. 2006. Genetic position of Valencia (Spain) in the Mediterranean basin according to *Alu* insertions. *American Journal of Human Biology*. 18: 187-195.
22. Kutuev I, Khusainova R, Karunas A, Yunusbayev B, Fedorova S, Lebedev Y, Hunsmann G, and Khusnutdinova E. 2006. From East to West: patterns of genetic diversity of populations living in four Eurasian regions. *Human Heredity*. 61: 1-9.
23. Chadli S, Wajih M, and Izaabel H. 2009. Analysis of *Alu* insertion polymorphism in South Morocco (Souss): use of markers in forensic science. *The Open Forensic Science Journal*. 2: 1-5.
24. Gomez-Perez L, Alfonso-Sanchez MA, Dipierri JE, Alfaro E, Garcia-Obregon S, De Pancorbo MM, Bailliet G, and Pena JA. 2011. Microevolutionary processes due to landscape features in the province of Jujuy (Argentina). *American Journal of Human Biology*. 23:177-184.
25. Kee BP, Chua KH, Lee PC, and Lian LH. 2012. Population data of six *Alu* insertions in indigenous groups from Sabah, Malaysia. *Annals of Human Biology*. 39: 505-510.
26. Hidalgo PC, Mut P, Ackermann E, Figueiro G, and Sans M. 2014. Questioning the "melting pot": analysis of *Alu* inserts in three population samples from Uruguay. *Human Biology*. 86: 83-92.
27. Santovito, A, Cervella P, and Delpero M. 2014. *Alu* insertion polymorphisms in four ethnic groups from northern Ivory Coast. *Anthropological Science*. 122: 37-43.
28. Swarovskaya MG, Marusin AV, Tacheeva TI, Khitrinskaya IY, Gafarov NI, and Stepanov VA. 2015. Analysis of eight polymorphic *Alu* elements in the Teleuts population. *Russian Journal of Genetics*. 51: 827-830.
29. Christie OO, Chosen OE, Samuel BJ, Favour CC, and Godwill EA. 2020. Evolutionary relationship of four major ethnic populations in Nigeria based on *Alu* PV92 insertion polymorphism. *International Journal of Human Genetics*. 20: 47-54.
30. Elwess NL, Duprey SL, Harney LA, Langman JE, Marino TC, Martinez C, McKeon LL, Moss CI, Myrie SS, and Taylor LR. 2008. *Alu* insertions and genetic diversity: a preliminary investigation by an undergraduate bioinformatics class. *Bioscene: Journal of College Biology Teaching*. 34: 34 -39.
31. Brooker RJ. 2015. Genetics: analysis and principles, 5th edition. McGraw-Hill Education, New York, NY, USA.
32. Charleston Southern University. 2021. A brief look at CSU. North Charleston, SC. [cited 13 June 2022]. Available from: www.charlestonsouthern.edu/about/at-a-glance/
33. Data USA. 2020. Enrollment. New York, NY. [cited 13 June 2022]. Available from: https://datausa.io/profile/university/charleston-southern-university
34. Dolan DNA Learning Center. 2006. Using an *Alu* insertion polymorphism to study human populations. Cold Spring Harbor Laboratory (New York). [cited 30 May 2022]. Available from: http://bioinformatics.dnalc.org/alu/animation/pdf/pv92.pdf
35. Lee PY, Costumbrado J, Hsu CY, and Kim YH. 2012. Agarose gel electrophoresis for the separation of DNA fragments. *Journal of Visualized Experiments*. 62: 1-5.
36. Deininger PL and Batzer MA. 1999. *Alu* repeats and human disease. *Molecular Genetics and Metabolism*. 67: 183-193.