U.S. DEPARTMENT OF JUSTICE
Office of Justice Programs

LSU PROPOSAL # 14120

# CATEGORICAL ASSISTANCE PROGRESS REPORT

FILE

Information provided will be used by the grantor agency to monitor grantee cash flow to ensure proper use of Federal funds. No further monies or other benefits may be paid out under this program unless this report is completed and filed as required by existing law and regulations (Uniform Administrative Requirements for Grants and Cooperative Agreements — 28 CFR, Part 66, Common Rule, and OMB Circular A-110).

| | | |
|---|---|---|
| 1. GRANTEE<br>Louisiana State University<br>Office of Sponsored Programs, 330 T. Boyd, | 2. AGENCY GRANT NUMBER<br>2001-IJ-CX-K004 | 3. REPORT NO.<br>5 |
| IMPLEMENTING SUBGRANTEE Baton Rouge, LA 70803 | 5. REPORTING PERIOD (Dates)<br>FROM: 07/01/2002 TO: ~~04/30/2003~~ 6/30/2003 | |
| SHORT TITLE OF PROJECT<br>LINE Elements | 7. GRANT AMOUNT<br>$370,513 | 8. TYPE OF REPORT<br>☐ REGULAR ☒ FINAL REPORT ☐ SPECIAL REQUEST |
| NAME AND TITLE OF PROJECT DIRECTOR<br>Mark Batzer, Ph.D. | 10. SIGNATURE OF PROJECT DIRECTOR<br>MAB | 11. DATE OF REPORT<br>05/26/2003 |

COMMENCE REPORT HERE (Continue on plain paper)

See the attached narrative.

RECEIVED
CONTROL DESK
OJP/OC

| | | |
|---|---|---|
| 13. CERTIFICATION BY GRANTEE (Official signature)<br>Karen Baicamonday Fron | JAMES L. BATES<br>DIRECTOR, OFFICE OF SPONSORED PROGRAMS | 14. DATE<br>5/26/03 |

OJP FORM 4587/1 (REV. 2-90)

**Final Progress Report 2001-IJ-CX-K004 05/01/01 – 04/30/03**

This is our final progress report from our NIJ grant 2001-IJ-CX-K004. The original objectives and aims of the grant are outlined below. The progress of the grant has been outstanding and we summarize the progress that we have made below. Seven original articles were published during the grant period and they are included as an appendix with this report. The abstracts of each of the articles are included in this progress report. At the end of the progress report we conclude with a section for future studies.

Objectives and Aims

The study of human population genetics and forensic DNA fingerprinting have been approached using a number of different nuclear and mitochondrial markers. The insertion of mobile elements into the human genome represents an alternative source of nuclear variability for forensic DNA profiling. The L1 family of long interspersed elements (LINEs) is present at a large copy number within the genome (100,000 copies/genome equivalent). L1 elements have been divided into subfamilies of related sequences based on nucleotide divergence. The L1 Hs (human-specific) Ta (transcriptionally active) LINE subfamily is the most recently formed group of LINEs, having amplified to about 4000 copies in the last 4-6 million years after the human/African ape divergence. Approximately 30% (1200) of the "young" Ta LINE elements appear to be polymorphic for presence/absence in the human genome. The distribution of the polymorphic LINE insertions is variable within diverse human population groups. Individuals that share polymorphic LINE elements inherited the alleles from a common ancestor (making the alleles identical by descent). This distinguishes LINE insertion polymorphisms from other traditional markers (e.g. restriction fragment length polymorphism and variable number of tandem repeats loci). These properties make the polymorphic LINE elements a unique class of nuclear markers for the geographic localization of unknown DNA samples and paternity testing. In this proposal we will identify "young" polymorphic Ta LINE elements and develop these elements as markers for forensic DNA profiling.

The specific aims of this project are:

1. The identification fifty new polymorphic LINE insertions.
2. Determination of the chromosomal location and phylogenetic distribution of each LINE element.
3. Determination of the human genetic variability associated with each polymorphic LINE repeat.
4. Develop multiplex polymerase chain reaction based assays and/or array based genotyping formats for these markers.

To summarize the progress that we have made we will provide point by point comments about our progress with respect to all the goals (listed together). Overall the grant resulted in five original journal articles as well as two review articles.

1) Identified over 140 new L1 insertion polymorphisms.
2) Determined the chromosomal location and phylogenetic distribution of each L1 element.
3) Determined the human genetic variation associated with the L1 insertion polymorphisms in major geographic groups.
4) Multiplex analysis of these loci was not reproducible as a result of the large PCR amplicons that contained a significant proportion (greater than 98%) of nearly identical genetic material (L1 element).

## ORIGINAL RESEARCH ARTICLES

Myers, J. S. *, B. J. Vincent*, H. Udall, W. S. Wakins, G. E. Kilroy, T. A. Morrish, G. D. Swergold, J. Henke, L. Henke, J. V. Moran, L. B. Jorde and M. A. Batzer (2002) *A comprehensive analysis of recently integrated human L1 Ta elements.* American Journal of Human Genetics **71:** 312-326.

*These authors contributed equally to this work.

## ABSTRACT

The Ta subfamily of Long INterspersed Elements (LINEs) is characterized by a 3 base pair "ACA" sequence in the 3' untranslated region and contains approximately 520 members in the

human genome. Here, we have extracted 468 L1Hs-Ta elements from the draft human genomic sequence and screened individual elements using polymerase chain reaction (PCR) assays to determine their phylogenetic origin and levels of human genomic diversity. One hundred twenty four of the elements amenable to complete sequence analysis were full-length (approximately 6 kb) and have apparently escaped any 5′ truncation. Forty-four of these full-length elements have two intact open reading frames and may be capable of retrotransposition. Sequence analysis of the Ta L1 elements showed a low level of nucleotide divergence with an estimated age of 1.99 million years old suggesting that expansion of the L1 Ta subfamily occurred after the divergence of humans and African apes. A total of 262 Ta L1 elements were screened with PCR based assays to determine their phylogenetic origin and the level of human genomic variation associated with each element. All of the Ta L1 elements analyzed by PCR were absent from the orthologous positions in non-human primate genomes except for a single element (L1HS72), that was also present in common (*Pan troglodytes*) and pygmy (*Pan paniscus*) chimpanzee genomes. Sequence analysis revealed that this single exception is the product of a gene conversion event involving an older pre-existing L1 element. One-hundred fifteen (45%) of the Ta L1 elements were polymorphic with respect to insertion presence or absence and will serve as identical by descent markers for the study of human evolution.

Romualdi, C., D. Balding, I. S. Nasidze, G. Risch, M. Robichaux, S. T. Sherry, M. Stoneking, M. A. Batzer and G. Barbujani (2002) *Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms.* Genome Research **12:** 602-612.

## ABSTRACT

Previous studies have reported that about 85% of human diversity at STR and RFLP autosomal loci is due to differences between individuals of the same population, whereas differences among continental groups account for only 10% of the overall genetic variance. These findings conflict with popular notions of distinct and relatively homogeneous human races, and may also call into question

the apparent usefulness of ethnic classification in, for example, medical diagnostics. Here, we present new data on 21 Alu insertions in 32 populations. We analyze these data along with three other large, globally dispersed datasets consisting of apparently neutral biallelic nuclear markers, as well as with a β-globin dataset possibly subject to selection. We confirm the previous results for the autosomal data, and find a higher diversity among continents for Y-chromosome loci. We also extend the analyses to address two questions: (1) whether differences between continental groups, although small, are nevertheless large enough to confidently assign individuals to their continent on the basis of their genotypes; (2) whether the observed genotypes naturally cluster into continental or population groups when the sample source location is ignored. Using a range of statistical methods we show that classification errors are at best around 30% for autosomal biallelic polymorphisms and 27% for the Y chromosome. Two datasets suggest the existence of 3 and 4 major groups of genotypes worldwide, respectively, and the two groupings are inconsistent. These results suggest that at random biallelic loci there is little evidence, if any, of a clear subdivision of humans into biologically-defined groups.

Roy-Engel, A. M. *, M. L. Carroll*, M. El-Sawy, A.-H. Salem, R. K. Garber, S. V. Nguyen, P. L. Deininger+, and M. A. Batzer+ (2002) *Non-traditional Alu evolution and primate genomic diversity.* Journal of Molecular Biology **316:** 1033-1040. [Cover Article]

*These authors contributed equally to this work.

+PLD and MAB are equal senior authors.

## ABSTRACT

Alu elements belonging to the previously identified "young" subfamilies are thought to have inserted in the human genome after the divergence of humans from non-human primates and therefore should not be present in non-human primate genomes. Polymerase chain reaction (PCR) based screening of over 500 Alu insertion loci resulted in the recovery of a few "young" Alu elements that also resided at orthologous positions in non-human primate genomes. Sequence analysis demonstrated these "young" Alu insertions represented gene conversion events of pre-existing ancient Alu elements or independent parallel insertions of older Alu elements in the same genomic

region. The level of gene conversion between Alu elements suggests that it has had a significant influence on the single nucleotide diversity within the genome. All the instances of multiple independent Alu insertions within the same small genomic regions were recovered from the owl monkey genome, indicating a higher Alu amplification rate in owl monkeys relative to many other primates. This study suggests that the majority of Alu insertions in primate genomes are the products of unique evolutionary events.

Kayser, M., S. Brauer, H. Schadlich, M. Prinz, M. A. Batzer, P. A. Zimmerman, B. A. Boatin and M. Stoneking (2003) *Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry.* Genome Research **13:** 624-634.

## ABSTRACT

The existence of significant geographic structure within U.S. ethnic populations is an important issue for forensic DNA scientists and for the medical genetics community. To investigate this issue, we analyzed 1705 haplotypes based on nine short tandem repeat (STR) loci on the Y-chromosome from nine to eleven groups each of African-Americans, European-Americans, and Hispanics. There were no significant differences in the distribution of Y-STR haplotypes amongst African-American groups, whereas European-American and Hispanic groups did exhibit significant geographic heterogeneity. However, for both of the latter groups the significant heterogeneity resulted from one sample; removal of that sample in each case eliminated the significant heterogeneity. Multidimensional scaling (MDS) analysis of $R_{ST}$ values indicated that African-American groups formed a distinct cluster, whereas there was some intermingling of European-American and Hispanic groups. By contrast, MDS analysis of published mtDNA haplotypes from largely the same groups (Melton et al. 2001) showed that African-Americans, European-Americans and Hispanic groups comprised three distinct clusters with no intermingling. Estimates of the amount of European-American admixture in African-American groups, based on two different methods, were 27.5-33.6%

for the Y-STR haplotypes and 9-15.4% for the mtDNA SSO-types; there was no significant heterogeneity in the amount of European-American admixture among groups. The lack of significant geographic heterogeneity among African-American, European-American, and Hispanic Y-STR and mtDNA haplotypes means that forensic DNA databases do not need to be constructed for separate geographic regions of the U.S. Moreover, absence of significant geographic heterogeneity for these two loci means that regional variation in disease susceptibility within ethnic groups is more likely to reflect cultural/environmental factors, rather than any underlying genetic heterogeneity.

Hedges, D. J., J. A. Walker, P. A. Callinan, J. G. Shewale, S. K. Sinha and M. A. Batzer (2003) *Mobile element-based assay for human gender determination.* Analytical Biochemistry **312:** 77-79.

## ABSTRACT

Determination of gender from human DNA samples is a common problem in forensics laboratories. While several PCR-based assays are currently available for human sex typing, each of the current approaches has limitations. Methods based on male-specific amplification, such as the amplification of the SRY locus, lack an internal positive control to discriminate between female DNA and male DNA which has failed to amplify for technical reasons. Restriction fragment length polymorphism (RFLP) assays based on sex-specific mutations at the ZFX/ZFY require a second enzyme digestion or hybridization step following the initial PCR amplification. A recent method proposed by Cali *et al.* based on a single adenine insertion within a tandem repeat array at the DXYS156 locus requires access to allele detection equipment potentially unavailable to forensics labs with limited resources. The most widely used approach is based on the *Amelogenin* locus, which yields different sized polymerase chain reaction (PCR) amplicons for the X and Y chromosome versions of the *Amelogenin* gene. However, this method misidentifies males as females in some cases due to a deletion in the *AMEL Y* region. This deletion has previously been reported to be

present at a frequency of 0.018% in Caucasian males, 1.85 % among Indians, and as high as 8% in Sri-Lankans. While the frequency of the deletion is relatively low, the crucial nature of forensic test results in circumstances such as rape and prenatal gender determination where there is risk for male-specific inherited disorders, makes any source of error a legitimate cause for concern. This has lead several researchers to recommend that *Amelogenin* not be relied upon as the sole determinant of gender. Here, we present an alternative PCR method of human gender identification based on the presence/absence of *Alu* sequences.

## REVIEW ARTICLES

Batzer, M. A. and P. L. Deininger (2002) *Alu repeats and human genomic diversity.* Nature Reviews Genetics **3:** 370-379.

## ABSTRACT

Alu elements have propagated to more than one million copies in primate genomes over 65 million years. This resulted in the generation of a series of Alu subfamilies that are of different ages. Alu elements impact the genome in a number of different ways including insertion mutations, recombination between elements, gene conversion, and gene expression. Alu insertion polymorphisms are a boon for the study of human population genetics and primate comparative genomics because they are neutral genetic markers of identical descent with known ancestral states.

Deininger, P. L. and M. A. Batzer (2002) *Mammalian retroelements.* Genome Research **12:** 1455-1465.

## ABSTRACT

The eukaryotic genome has undergone a series of epidemics of amplification of mobile elements that have resulted in most eukaryotic genomes containing much more of this 'junk' DNA than actual coding DNA. The majority of these elements utilize an RNA intermediate and are termed retroelements. Most of these retroelements appear to amplify in evolutionary waves that insert in the genome and then gradually diverge. In humans, almost half of the genome is recognizably derived from retroelements, with the two elements that are currently actively amplifying, L1 and Alu, making up about 25% of the genome and contributing extensively to disease. The mechanisms of this amplification process are beginning to be understood, although there are still more questions than answers. Insertion of new retroelements may directly damage the genome, and the presence of multiple copies of these elements throughout the genome has longer-term influences on recombination events in the genome and more subtle influences on gene expression.

## Students and Postdocs supported by the grant

| Student | Degree | Present Location |
|---|---|---|
| Jeremy Myers | Ph.D. | Postdoc, Vanderbilt University, Nashville, TN |
| Bethaney Vincent | Ph.D. | Medical School, Vanderbilt University, Nashville, TN |
| Marion Carroll | Ph.D. | Assistant Professor, Xavier University, New Orleans, LA |
| Dale Hedges | Ph.D. | In progress, Louisiana State University, Baton Rouge, LA |
| Pauline Callinan | Ph.D. | In progress, Louisiana State University, Baton Rouge, LA |
| Gregory Risch | M.S. | Alabama State Police Lab., Birmingham Alabama |
| Steve Sherry | Post Doc | Staff Scientist, Natl Cent for Biotechnology Washington, DC |
| Abdel-Halim Salem | Post Doc | Dept. of Anatomy, Suez Canal University |

## Future Studies and Conclusions

It is clear that the identical by descent mobile element insertion polymorphisms will play a useful role in paternity/identity testing and in providing the geographic affiliations of unknown human DNA samples. These elements are a unique source of neutral population specific alleles within the human genome. However, each of these genetic systems contains only two alleles so a number of markers are required for the accurate assignment of the affiliation of unknown DNA samples, and each of the markers must be characterized in as many geographically distinct populations as possible. Therefore, further research providing a high resolution map of the genetic variation from each of the newly identified L1 insertion polymorphisms will be required to effectively utilize them for them for geographic affiliation studies. Mobile elements also represent a novel source of genetic variation for the development of species specific PCR amplicons for the identification and quantitation of DNA. We have developed an assay for human sex typing based upon mobile element insertion polymorphisms.