Open Access Journal | Journal of Progressive Research in Biology (JPRB)

ISSN 2454-1672

SCITECH RESEARCH ORGANISATION

Volume 1, Issue 1

Published online on April 29, 2015

Journal of Progressive Research in Biology

www.scitecresearch.com/journals

Are the Available DNA Sequence Alignments and Search Tools Reliable for Alu Research?

Dr. Sridhar Ramachandran.

Department of Informatics, Indiana University SE, New Albany, USA.

Abstract

The results of most bioinformatics and microbiology research studies that use SINEs depend entirely on the mechanism used to detect and count these element instances in the genome. Hence, the reliability and accuracy of the DNA sequence alignment and search tool is vital for genetic researches. This research report presents the findings from testing the reliability of some popular DNA Sequence alignment and search (SA&S) tools using a test/known genomic sequence. The findings reveal the need for novel tool design.

Keywords: Sequence Alignment; DNA Search Tools; Alu Polymorphism.

Introduction

It has been known for some time that organisms exhibit large variations in genome sizes which do not correlate with organismic complexity [8]. Much of this variation can be traced to non-coding DNA, which in many organisms is present in vast excess over coding DNA. Approximately 98% of the human genome is made up of regions that do not code for proteins [9]. These non-transcribed sequences, or "Junk DNA", are widely believed to consist largely of useless DNA leftovers from past evolutionary permutations [15]. However, this so-called Junk DNA is far from useless to genomic researchers and bioinformaticians.

A significant proportion of the Junk DNA is comprised of repetitive sequences. A major category of Junk repetitive sequences within all mammalian genomes studied to date is the Short Interspersed Nuclear Elements (SINEs) that account for as much as 10% of all genomic sequence. SINE elements are genomic hitchhikers [11] and move within the genome by either DNA or RNA mediated duplication events [10]. Within the human genome, there are approximately one million copies of the Alu family of SINEs alone. Alus require forming of an RNA transcript that must then be reverse transcribed and inserted into a new location in the genome [14]. Thus Alus are believed to have colonized the genome by a 'copy and paste' mechanism [7] and have actively copied and pasted themselves in the genome at different time periods. However, the means by which Alus have reached their current high genomic abundance remains unclear.

Proliferation of Alus in a host is a unidirectional process, whereby inserted copies of distinct elements are not precisely removed, but remain and decay over time because of random mutation [12]. Most Alus insert innocuously into nonfunctional regions and can provide an excellent record of biological history that is largely free from character reversals and parallel evolution [1]. These characteristics of Alus make them extremely useful tools for characterizing the genome.

Alu elements are generally detected using DNA sequence alignment and search tools. The results of studies using Alu elements mostly depend on the mechanism used to detect and count Alu instances. Hence, the reliability and accuracy of the DNA sequence alignment and search tool is vital for genetic researches that use Alus.

Research Report

The Alu repeats are divided into three broad sub-families based on their evolutionary age. Subfamily AluJ is believed to be the oldest, subfamily AluS being the intermediate and AluY subfamily being the youngest [6]. These Alu elements that amplified at different stages of the primate evolution have key diagnostic differences that allow them to be classified into subfamilies [3] [2]. The time line for the different Alu subfamilies is shown in **Figure 1.**

ISSN 2454-1672

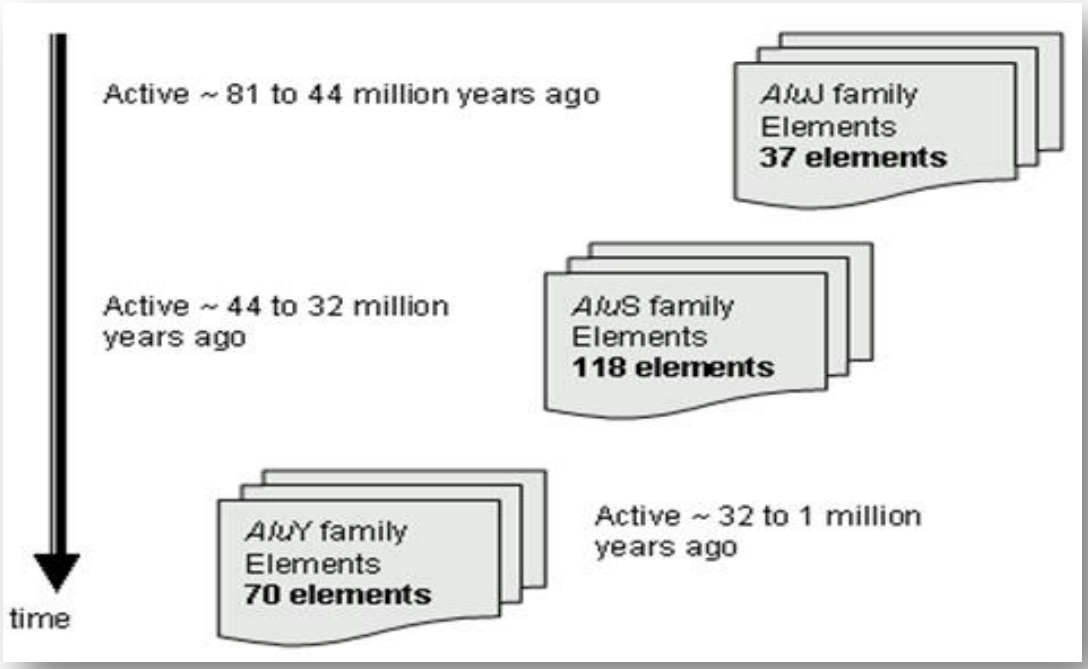

Figure 1: Alu Subfamilies

To test the reliability of DNA Sequence alignment and search (SA&S) tools, a synthetic genome with various Alu insertion polymorphisms including the Alu-within-Alu polymorphism was prepared (see **Figure 2**).

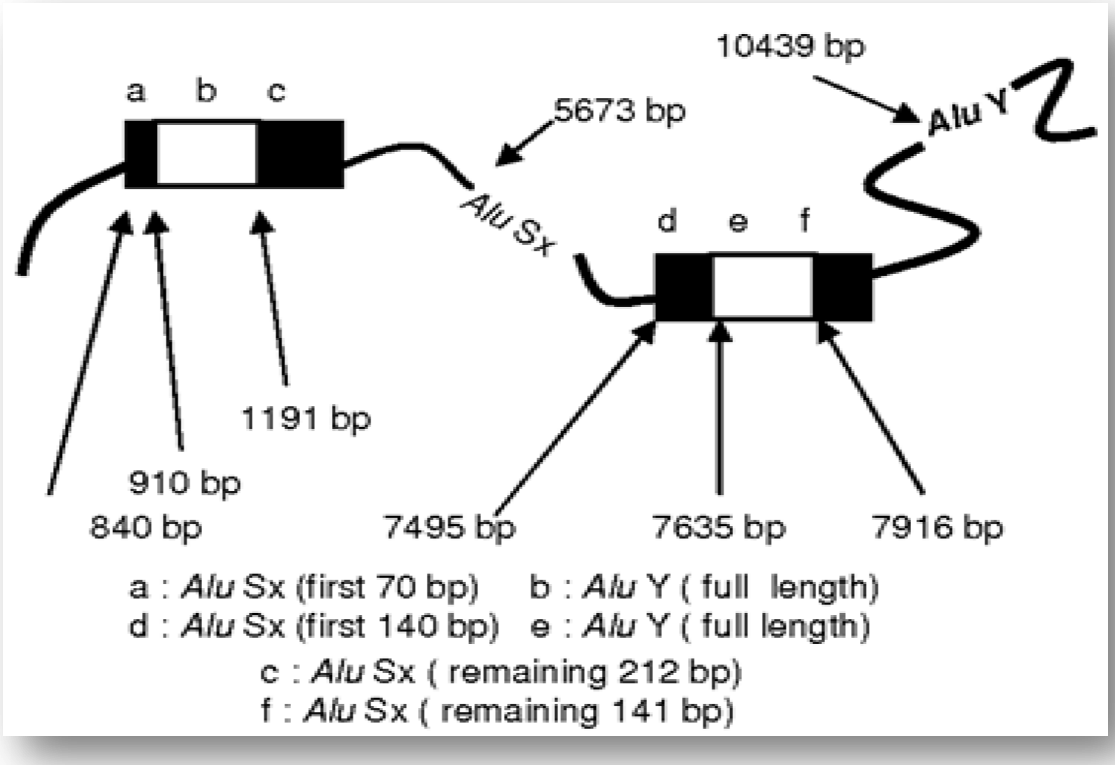

Figure 2: Synthetic Genome for Testing

ISSN 2454-1672

Using the synthetic genome shown in **Figure 2,** eight popular DNA sequence alignment and search tool were tested. The results of the test are shown in **Table 1.**

Table 1: Results of the Test

SA&S tool	Test with the synthetic genome	Finding
EST2GENOME	Only detected 2/8 *Alus*	Unreliable
NEEDLE	Only detected 2/8 *Alus*	Unreliable
NEEDLEALL	Only detected 2/8 *Alus*	Unreliable
SUPERMATCHER	Only detected 2/8 *Alus*	Unreliable
WORDFINDER	Only detected 2/8 *Alus*	Unreliable
Jalinger	Only detected 2/8 *Alus*	Unreliable
zPicture	Only detected 4/8 *Alus*	Unreliable
LALIGN/PLALIGN	Only detected 4/8 *Alus*	Unreliable

Discussion

When the synthetic genome (Figure 2) was tested on some popular DNA SA&S tools (shown in Table 1), it was found that individual and complete element insertion events were perfectly recognized by the tools. However when Alu-within-Alu events were inserted into the synthetic genome, the tool failed to report an accurate count of the number of Alu insertion events. This experiment thus identifies that different Alu insertion polymorphisms can affect the count of Alu events reported by search tools. This preliminary observation itself puts all researches that have used the Alus identified using DNA tools that we tested into question.

Alus are believed to prefer sites that are locally rich in A+T nucleotides [4]. The oligo-dA-rich (poly (A)) tails and middle (A) rich regions of Alu elements have previously been shown to serve as nuclei for the genesis of simple sequence repeats [4]. Alus are known to preferentially insert into the A tail of other Alus and thus are often found clustered adjacent to existing Alu elements [13]. The presence of two ‘A’ rich regions within the Alu element (in the middle and in the poly (A) tail) could increase the likelihood that one Alu element may insert within another [13,5]. Unfortunately, the popular DNA SA&S tools do not accommodate genetic polymorphism like Alu-within-Alu knowledge into their algorithm design and the findings from this research throws open the need for novel tool design.

Acknowledgments

My thanks to my undergraduate research student Mr. Kelley for helping retest my findings thus doubly verifying the results.

References

[1] A.M. Shedlock, K. Takahashi and N. Okada, "SINEs of speciation: tracking lineages with retroposons", Trends in Ecology and Evolution, vol.19 (10), October 2004, pp. 545 - 553.

[2] A.L. Price, E. Eskin, and P.A. Pevzner, “Whole-genome analysis of Alu repeat elements reveals complex evolutionary history”, Genome Re-search, vol. 14, November 2004, pp. 2245 – 2252.

[3] A.M. Roy-Engel, M.L. Carroll, M. El-Sawy, A.H. Salem, R.K. Garber, S.V. Nguyen, P.L. Deininger, M.A. Batzer, “ Non- traditional Alu evolution and primate genomic diversity”, J. Mol. Biol., vol. 316, no. 5, March 8, 2002, pp. 1033 – 1040.

[4] A-H. Salem, G. E. Kilroy, W. S. Watkins, L. B. Jorde, and M. A. Batzer, "Recently Integrated Alu Elements and Human Genomic Diversity", Molecular Biology. Evolution, vol. 20(8), pp. 1349–1361, 2003.

[5] D. Comas, S. Plaza, F. Calafell, A. Sajantila and J. Bertranpetit, “Recent Insertion of an Alu Element within a Polymorphic Human-Specific Alu Insertion,” Molecular Biology and Evolution, vol. 18, pp. 85-88, 2001.

ISSN 2454-1672

[6] D. Grover, P.P. Majumder, C.B. Rao, S.K. Brahmachari, and M. Mukerji, “Nonrandom Distribution of Alu Elements in Genes of Various Functional Categories: Insight from Analysis of Human Chromosomes 21 and 22”, Molecular Biology and Evolution, vol. 20(9), pp. 1420 – 1424, 2003.

[7] D.J. Hedges, P.A. Callinan, R. Cordaux, J. Xing, E. Barnes, and M.A. Batzer, “ Differential Alu Mobilization and Polymorphism Among the Human and Chimpanzee Lineages”, Genome Research, vol. 14, 2004, pp. 1068 – 1075.

[8] G. S. Attard, A. C. Hurworth and J. P. Jack, "Language-like features in DNA: transposable element foot-prints in the genome", Europhys. Lett, 36 (5), 1996, pp. 391-396.

[9] J. Brosius, "How significant is 98.5% ‘junk’ in mammalian genomes?", Bioinformatics, vol. 19, Suppl. 2, 2003, pp. ii35.

[10] J. Gephart, “Tracing Human Origins with “Alu” Markers”, http://www. pnl.gov/er_news/10 _95/ er_news/XALU-MARK.JG.html, March 20, 2006.

[11] J. Whitfield, “How the sequence got the way it is”, news@nature.com, http://news .nature .com//-news/2001 /010215/010215-1.html, February 12, 2001.

[12] K. Han, J. Xing, H. Wang, D.J. Hedges, R. K. Garber, R. Cordaux, and M.A. Batzer, "Under the genomic radar: The Stealth model of Alu amplification", Genome Research, vol.5 , 2005, pp. 655 – 664.

[13] M. El-Sawy, and P. Deininger, “Tandem insertions of Alu elements”, Cytogenetic and Genome Research, vol. 108, pp. 58 – 62, 2005.

[14] P.L. Deininger, and M.A. Batzer, “Mammalian Retroelements”, Genome Research, vol. 12, 2002, pp. 1455 – 1465.

[15] W. Makalowski, “Not Junk After All”, Science, vol. 300, May 23, 2003, pp. 1246 – 1247.