Gene 561 (2015) 283–291

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Genome-wide target site triplication of *Alu* elements in the human genome

Wooseok Lee [a], Seyoung Mun [a,b], Keunsoo Kang [c], Lothar Hennighausen [b], Kyudong Han [a,d,*]

[a] Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan 330-714, Republic of Korea
[b] Laboratory of Genetics and Physiology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892, USA
[c] Department of Microbiology, Dankook University, Cheonan 330-714, Republic of Korea
[d] DKU-Theragen institute for NGS analysis (DTiNa), Cheonan 330-714, Republic of Korea

ARTICLE INFO

Article history:
Received 19 December 2014
Received in revised form 2 February 2015
Accepted 15 February 2015
Available online 19 February 2015

Keywords:
Target site
Alu
Mobile element
Retrotransposon
TSD-TSD recombination

ABSTRACT

Alu elements are the most successful short interspersed elements in primate genomes and their retrotransposition is a major source of genomic expansion. *Alu* elements integrate into genomic regions through target-site primed reverse transcription, which generates target site duplications (TSDs). Unexpectedly, we have identified target site triplications (TSTs) at some loci, where two *Alu* elements in tandem share one direct repeat. Thus, the three copies of the repeat are present. We located 212 TST loci in the human genome and examined 25 putative human-specific TST loci using PCR validation. As a result, 12 human-specific TST loci were identified. These findings suggest that unequal homologous recombination between TSDs can lead to TST. Through this mechanism, the copy number of *Alu* elements could have increased in primate genomes without new *Alu* retrotransposition events. This study provides new insight into the augmentation of *Alu* elements in the primate genome.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Alu elements are non-autonomous retrotransposons with a size of approximately 300 nucleotides. They are derived from 7SL RNA prior to the primate radiation (Ullu and Tschudi, 1984) and have successfully expanded, reaching more than 1.2 million copies in the human genome (Han et al., 2007). *Alu* elements normally propagate through an RNA intermediate in a process known as retrotransposition, which is thought to occur by a mechanism of target-primed reverse transcription (TPRT). In general, novel *Alu* retrotransposition produces "target-site duplications" (TSDs), which are direct repeats (7–20 bp) and flank the element on both sides (Jurka, 1997; Cordaux and Batzer, 2009).

To date, much research has been conducted to explain the successful amplification of *Alu* elements in the human genome (Kim et al., 2012), and the availability of the chimpanzee reference genome has accelerated the research (Chimpanzee and Analysis, 2005). Despite those efforts, the mechanism yielding the expansion of *Alu* elements is not fully understood. Currently, the master gene model is generally accepted to explain their amplification in the host genome. It has been proposed that most of the *Alu* elements have been amplified through a small number of "master genes" featuring long-lived high activity (Shen et al., 1991). Each of the master genes is capable of generating a novel *Alu* subfamily. Under the master gene model, relatively old *Alu* subfamilies, however, have accumulated many base substitutions resulting in a decrease of *Alu* amplification rate and an eventual loss of proliferation. Thus, another model, "stealth model" was proposed to accommodate the expansion and survival of *Alu* elements in the human genome. This model suggests that low-activity *Alu* elements maintain their retrotransposition capacity over extended periods of time. In contrast to master genes, stealth drivers are not responsible for generating the majority of new *Alu* copies but they are capable to produce the master genes in suitable genomic environment (Han et al., 2005b).

The number of *Alu* elements in the human genome has increased through their retrotransposition. Here, we introduce a novel mechanism, TSD (target site duplication)–TSD recombination, that ensures the increase of *Alu* elements, without their retrotransposition. Each *Alu* element is flanked by two identical DNA segments, TSDs and the sequence identity between them promotes unequal homologous recombination, leading to either the duplication of an *Alu* element with target site triplication (TST) or the precise deletion of an *Alu* element in other allele (Kazazian, 2004; van de Lagemaat et al., 2005). We computationally identified TSD–TSD recombination candidates in the human genome and inspected them based on the presence of TST and the diversity between two contiguous *Alu* elements. We

Abbreviations: TSD, target site duplication; TST, target site triplication; PCR, polymerase chain reaction; INDEL, insertion/deletion; ARMD, *Alu* recombination-mediated deletion; TPRT, Target-primed reverse transcription.

* Corresponding author at: Department of Nanobiomedical Science & BK21 PLUS NBM Global Research Center for Regenerative Medicine, Dankook University, Cheonan, Republic of Korea.
E-mail address: jim97@dankook.ac.kr (K. Han).

http://dx.doi.org/10.1016/j.gene.2015.02.052
0378-1119/© 2015 Elsevier B.V. All rights reserved.

characterized the impact of TSD–TSD recombination in primate genome evolution by investigating their genomic environments and further testified human-specific TST events with experimentation.

2. Materials and methods

2.1. Computational data mining and manual inspection

To identify TST candidates in the human genome (hg38), the following steps were conducted using custom Perl scripts. First, RepeatMasker track was downloaded from the UCSC genome browser via the table browser (http://genome.ucsc.edu/) and genome coordinates of *Alu* elements were extracted. Second, consecutive *Alu* elements (length difference <80 bp) in proximity (<50 bp distance) with same direction and subfamily were obtained. Third, upstream sequence of the first element was aligned to the downstream sequence of the second element in each pair by using the BLASTN algorithm (Altschul et al., 1990) and non-aligned candidates were discarded. To determine whether these candidates were generated by the TSD–TSD recombination event, we reanalyzed 400 bp flanking sequences of the candidates with the RepeatMasker program and manually filtered out the output. A total of 6938 TST candidates were identified.

2.2. PCR and DNA sequencing analysis

The TSD–TSD recombination candidates, TST candidate, were verified through PCR. For the PCR, we used four different genomic DNAs of human, chimpanzee, gorilla and orangutan as PCR templates. The cell lines used to isolate the genomic DNAs were as follows: *Homo sapiens* (human; NA10851, Coriell Cell Repository, Camden, NJ), *Pan troglodytes* (common chimpanzee), *Gorilla gorilla* (gorilla) and *Pongo pygmaeus* (Bornean orangutan). To evaluate polymorphism rates of human-specific TST loci, we amplified them through PCR with a human population panel composed of 80 unrelated individuals (20 from each of four populations: African American, South American, European, and Asian), that was available from previous studies in our lab.

Oligonucleotide primers for each TST locus were designed using Primer3 (http://bioinfo.ut.ee/primer3-0.4.0) and then computationally tested utilizing both the Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html) and UCSC's In-Silico PCR (http://genome.ucsc.edu/cgi-bin/hgPcr?command=start). Each PCR was conducted in 25 μl reactions with 15–50 ng DNA, 10 μM of each oligonucleotide primer, and 10 μl of PCR master mix containing DNA polymerase (2× EF Taq PCR Smart mix 4, SolGent, Korea). The conditions for the PCR were an initial denaturation step of 5 min at 95 °C, followed by 35 cycles of PCR at 20 s of denaturation at 95 °C, 40 s at the annealing temperature, and ≤4 min of extension at 72 °C, followed by a final extension step of 5 min at 72 °C. The PCR products were loaded on 1–4% agarose gels, depending on the PCR product sizes, stained with ethidium bromide, and visualized with UV fluorescence (Bio-Rad, Hercules, CA). In cases where the expected size of the PCR product was greater than 2 kb, KOD FX (Toyobo, Japan) was used, following the manufacturer's instructions. The information of primer pairs was described in Supplementary Table 1.

When necessary, individual PCR products were purified from the agarose gels using the FavorPrep GEL/PCR Purification Mini Kit (FAVORGEN, Taiwan) and cloned using TOPO-TA Cloning kit (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. For each sample, at least three colonies were randomly selected and subject to colony PCR to be sequenced. DNA sequencings of the samples were performed on an Applied Biosystems ABI3130*XL* Genetic Analyzer (Applied Biosystems, Foster City, CA), using dideoxy chain-termination sequencing. Raw sequence data were analyzed using BioEdit program in the version 7.2.1 software package (http://www.mbio.ncsu.edu/BioEdit/BioEdit.html).

2.3. Analysis of flanking sequences

For each TST locus, 5′- and 3′-flanking sequences (each 10 kb) were collected. The GC content of the flanking sequences from each TST locus was calculated by using the combined 20 kb of flanking sequences via in-house Perl script, which excluded N_s from the analysis. For the gene density analysis, we estimated using the NCBI Map Viewer utility,

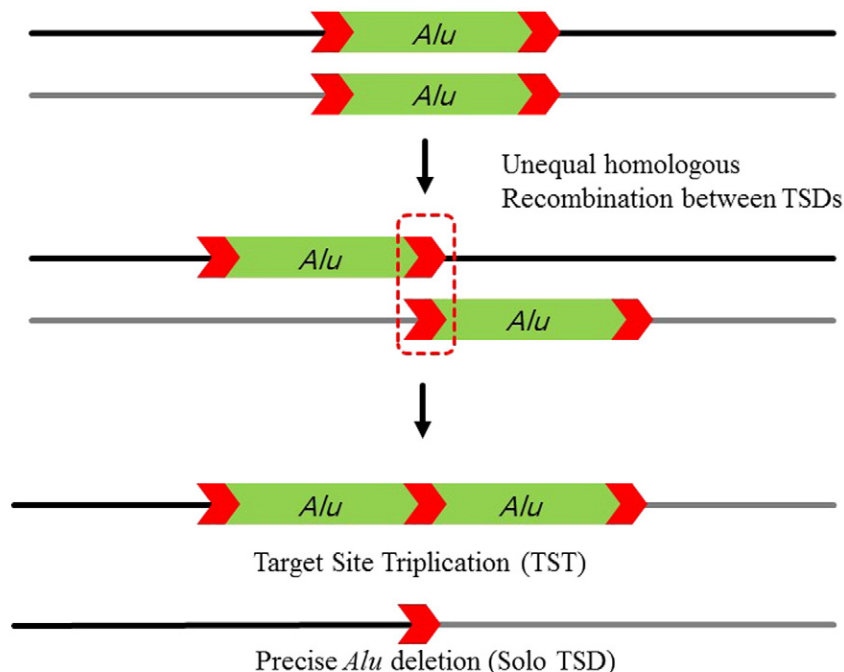

Fig. 1. The mechanism of target site triplication. Each *Alu* elements are flanked by two identical DNA segments, TSDs. The sequence identity between the two segments causes unequal homologous recombination between them, leading to the duplication of an *Alu* element and target site triplication (TST) in the human genome.

Download English Version:

https://daneshyari.com/en/article/2815820

Download Persian Version:

https://daneshyari.com/article/2815820

Daneshyari.com