Available online at www.sciencedirect.com
ScienceDirect
Gene 390 (2007) 18–27
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons

Jungnam Lee [a,1], Richard Cordaux [a,1], Kyudong Han [a,1], Jianxin Wang [b], Dale J. Hedges [a], Ping Liang [b,2], Mark A. Batzer [a,*,2]

[a] *Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA*
[b] *Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA*

Received 24 April 2006; received in revised form 5 August 2006; accepted 25 August 2006
Available online 16 September 2006

Abstract

The long interspersed element-1 (LINE-1 or L1) is a highly successful retrotransposon in mammals. L1 elements have continued to actively propagate subsequent to the human–chimpanzee divergence, ~6 million years ago, resulting in species-specific inserts. Here, we report a detailed characterization of chimpanzee-specific L1 subfamily diversity and a comparison with their human-specific counterparts. Our results indicate that L1 elements have experienced different evolutionary fates in humans and chimpanzees within the past ~6 million years. Although the species-specific L1 copy numbers are on the same order in both species (1200–2000 copies), the number of retrotransposition-competent elements appears to be much higher in the human genome than in the chimpanzee genome. Also, while human L1 subfamilies belong to the same lineage, we identified two lineages of recently integrated L1 subfamilies in the chimpanzee genome. The two lineages seem to have coexisted for several million years, but only one shows evidence of expansion within the past three million years. These differential evolutionary paths may be the result of random variation, or the product of competition between L1 subfamily lineages. Our results suggest that the coexistence of several L1 subfamily lineages within a species may be resolved in a very short evolutionary period of time, perhaps in just a few million years. Therefore, the chimpanzee genome constitutes an excellent model in which to analyze the evolutionary dynamics of L1 retrotransposons.

Keywords: L1 elements; Retrotransposons; Human; Chimpanzee; Species-specific; Polymorphism

1. Introduction

Long interspersed elements-1 (LINE-1 or L1) are the most successful autonomous retrotransposons in mammals. A full-length functional L1 element is about 6 kb in length and contains a 5′ untranslated region (UTR) bearing an internal RNA polymerase II promoter, two non-overlapping open reading frames (ORF1 and ORF2), which are separated by an ~60-bp-long intergenic spacer, and a 3′ UTR ending in a poly(A) tail (Kazazian and Moran, 1998). ORF1 encodes an RNA-binding protein that has nucleic acid chaperone activity *in vitro*, and ORF2 encodes both reverse transcriptase and endonuclease activities (Mathias et al., 1991; Feng et al., 1996; Kolosha and Martin, 1997). L1 elements propagate through an RNA intermediate in a process known as retrotransposition, which is thought to occur by a mechanism termed target primed reverse transcription; the insertion process typically results in 7–20-bp-long target site duplications flanking each side of the L1 element (Fanning and Singer, 1987; Luan et al., 1993).

With >500,000 copies, L1 elements account for ~17% of the human genome (Lander et al., 2001). The L1 family emerged around 120 million years (myrs) ago (Smit et al., 1995; Khan et al., 2006) and is still actively expanding in humans, as demonstrated by the existence of highly polymorphic L1 elements in human populations (Sheen et al., 2000; Myers et al., 2002;

Abbreviations: LINE-1 or L1, long interspersed element-1; UTR, untranslated region; ORF, open reading frame; Pt, *Pan troglodytes*-specific; Hs, *Homo sapiens*-specific; myrs, million years.

* Corresponding author. Tel.: +1 225 578 7102; fax: +1 225 578 7113.
E-mail address: mbatzer@lsu.edu (M.A. Batzer).

[1] These authors contributed equally to this research.
[2] These authors are equal senior authors.

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.08.029

Badge et al., 2003; Boissinot et al., 2004; Seleme et al., 2006; Wang et al., 2006) and *de novo* L1 insertions responsible for genetic disorders (Chen et al., 2005). The detection of several hundred species-specific L1 insertions in both the human and chimpanzee genomes further supports the recent mobilization of this family of retrotransposons (Mathews et al., 2003; CSAC, 2005; Mills et al., 2006). Contrary to the non-autonomous *Alu* retrotransposons in which different subfamilies are capable of concomitant expansions (Batzer and Deininger, 2002; Xing et al., 2004; Hedges et al., 2005), a single line of successive L1 subfamilies has amplified within the past 40 myrs in the primate lineage leading to humans (Khan et al., 2006). L1 subfamilies can be distinguished by diagnostic substitutions that are shared by all members of any given subfamily. For example, five subfamilies are thought to have amplified in hominoid primates (*i.e.* humans and apes) within the past 25 myrs, named L1PA1 to L1PA5 (Smit et al., 1995; Boissinot et al., 2000; Lander et al., 2001; Khan et al., 2006). The most recently evolved, *Homo sapiens*-specific (Hs) L1 subfamilies have been well characterized (Boissinot et al., 2000; Myers et al., 2002; Ovchinnikov et al., 2002; Salem et al., 2003a; Boissinot et al., 2004) and the recent completion of the chimpanzee genome sequence (CSAC, 2005) facilitates comparisons of the recent patterns of diversity and evolution of L1 subfamilies since the divergence of human and chimpanzee, ~6 million years ago (Goodman et al., 1998). Global overviews of Hs and *Pan troglodytes*-specific (Pt) L1 elements have previously been published (CSAC, 2005; Mills et al., 2006). Here, we report a detailed characterization of Pt L1 subfamily diversity and a comparison with their Hs counterparts. Our results indicate that L1 elements have experienced drastically different evolutionary fates in humans and chimpanzees within the past ~6 myrs.

2. Materials and methods

2.1. Computational identification of L1 elements

We identified all L1 elements with complete 3′ end sequences in the human genome (hg16, UCSC July 2003 freeze) by Basic Local Alignment Search Tool (BLAST) querying the genome with the 3′-most 50 bp preceding the poly-A tail of the L1 consensus sequence. This strategy yielded ~110,000 candidate elements, corresponding to the most recent fraction of all L1 elements inserted in the human genome. Next, 300-bp-long sequences covering each L1 3′-end and 100 bp of flanking sequence immediately downstream the poly-A tail were extracted. The exact terminus of the poly-A tails in these L1 sequences was determined by a BLAST search with the 50-bp L1 consensus sequence to which a tract of 100 adenosines was added. The sequences were used as queries for BLAST searches against the chimpanzee genome sequence (UCSC Nov. 2003 freeze). Queries with matches limited to the 100-bp L1 3′ end flanking regions in human were collected as candidates representing the orthologous pre-integration sites of the human L1 insertions. Then, we extracted the 800-bp region centered at the chimpanzee pre-integration site, along with the human L1 insertion and 400-bp upstream and downstream flanking sequence. To reduce false positives, pairs of chimpanzee and human non-L1 genomic sequences were required to exhibit >95% identity over their entire length. This resulted in 1989 candidate Hs L1 insertions. The procedure was repeated by reversing the order of the human and chimpanzee genome sequences to identify candidate Pt L1 insertions, resulting in the recovery of 1207 loci. All candidate loci were subsequently subjected to manual verification, yielding a total of 1835 Hs and 1190 Pt L1 elements.

2.2. PCR amplification and DNA sequencing

Cell lines used to isolate DNA samples were as follows: human (*H. sapiens*) HeLa (American Type Culture Collection [ATCC] number CCL2), common chimpanzee Clint (*P. troglodytes*; cell line NS06006B), gorilla (*Gorilla gorilla*; cell line AG05251) and orangutan (*Pongo pygmaeus*; cell line ATCC CR6301). DNA samples from 20 European, 20 African American and 20 Asian human individuals isolated from peripheral blood lymphocytes were available from previous studies in our lab, and DNA samples from 20 South American individuals were obtained from the Coriell Institute for Medical Research. A common chimpanzee (*P. troglodytes*) population panel composed of 12 unrelated individuals of unknown geographic origin was obtained from the Southwest Foundation for Biomedical Research.

Oligonucleotide primers for the PCR amplification of L1 elements were designed using the software Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). PCR amplification of each locus was performed in 25 μl reactions using 10–50 ng DNA, 200 nM of each oligonucleotide primer, 200 μM dNTPs in 50 mM KCl, 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.4) and 2.5 U Taq DNA polymerase. Each sample was subjected to an initial denaturation step of 5 min at 95 °C, followed by 35 cycles of PCR at 1 min of denaturation at 95 °C, 1 min at the annealing temperature, 1 min of extension at 72 °C, followed by a final extension step of 10 min at 72 °C. The resulting products were loaded on 2% agarose gels, stained with ethidium bromide, and visualized using UV fluorescence. Detailed conditions for all PCR assays designed in this study are available in Supplemental Table 1.

Individual PCR products were purified from the gels using the Wizard® gel purification kit (Promega) and cloned into vectors using the TOPO-TA Cloning® kit (Invitrogen), according to the manufacturer's instructions. DNA sequencing was performed using chain termination sequencing on an Applied Biosystems 3100 automated DNA sequencer. The DNA sequences from this study have been deposited in GenBank under accession numbers DQ375560–DQ375750.

PCR amplification of 5 full-length L1 loci was performed in 50 μl reactions using 200 ng DNA, 300 nM of each oligonucleotide primer, 200 μM dNTPs, 1 mM $MgSO_4$, 2% DMSO, and 2 U KOD Hifi DNA polymerase (Novagen). Each sample was subjected to heating for 2 min at 94 °C to activate the polymerase, followed by 35 cycles of PCR at 15 s of denaturation at 94 °C, 30 s of annealing at 60 °C, and 5 min of extension at 72 °C. The PCR products were purified using the