Genomics 94 (2009) 278–283

Contents lists available at ScienceDirect

Genomics

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

Short Communication

Discovery of a long inverted repeat in human POTE genes

Yong Wang, Frederick C.C. Leung *

School of Biological Sciences and Genome Research Centre, The University of Hong Kong, Pokfulam, Hong Kong, China

ARTICLE INFO

Article history:
Received 16 February 2009
Accepted 13 May 2009
Available online 20 May 2009

Keywords:
POTE gene
Intronic LIR
Stem-loop
Phylogeny
Primates

ABSTRACT

POTE gene family is tightly related to prostate, ovary, testis and placenta cancers. We recently identified an intronic long inverted repeat (LIR) in some members of the POTE gene family. Due to the capacity of inducing gene amplification, the POTE intronic LIRs may be involved in over-expression of the POTE genes. Our study aimed to understand the origin of the LIR in primates. We collected the LIR and its flanking sequences within rhesus monkey, chimpanzee and human genomes. The rhesus monkey genome only has half-sized LIRs (lack one repeat copy), whereas the human and chimpanzee genomes contain both full-sized and half-sized LIRs. Phylogenetic tree indicates that the LIR is formed after divergence of rhesus monkey and the common ancestor of human and chimpanzee. The POTE genes containing a full-sized LIR were amplified in the human genome.

Introduction

Recently, the structure and function of a primate-specific gene family POTE (expressed in prostate, ovary, testis, and placenta) are reported [1,2]. In humans, it consists of 11 highly homologous members, located on 7 different chromosomes [2]. The gene family attracted much attention because the novel genes are not only specifically expressed in normal prostate, ovary, testis and placenta, but also in many cancer tissues [1,3]. In other normal tissues, the expression of the POTE genes is undetectable, enabling them to be subjected to tests for immunotherapy of prostate cancers [3]. The gene members can be classified into: group 1 (POTE 8α, 8β), group 2 (POTE 15, 18 and 21) and group 3 (POTE 2α, 2β, 2β', 2γ, 2δ, 14α, 14β and 22), with respect to similarity [2,4]. The proteins encoded by the gene family are in the range of 32 to 80 kDa [2]. They are found on the inner aspect of plasma membrane and suggested to act in signaling pathway in the reproductive system [2]. Probably due to the loss of stop codons, POTE 2α, 2β, and 2γ encode chimeric POTE–actin, a fusion protein of POTE and actin [5].

Although the gene family was well characterized, the variance in introns had not yet been surveyed. In our previous work, a long inverted repeat (LIR) in the intron between exon 10 and exon 11 was identified [6]. The full-sized structure of the LIR only presents in group 3 of the POTE gene family, and the members in other groups lack one arm (one repeat copy of the LIR) of the LIR. This finding suggests its potential causal role in making the variants within the gene family.

In this study, we performed bioinformatic tests aiming to compare the LIRs in primates. We collected the LIRs (no matter full or half-sized) and their flanking sequences in human, chimpanzee (*Pan troglodytes*) and rhesus monkey (*Macaca mulatta*) genomes. All the collected sequences in rhesus monkey genome have only one arm of the LIR. Phylogeny of the sequences indicates that the ancestral sequence resides on chromosome 8 and the full structure of the LIR was developed in POTE genes belonging to the group 3 in the common ancestor of chimpanzee and human.

Results

POTE22 gene was found in the pericentromeric region of human chromosome 22. An LIR was located in the last intron, 163 bp to the last exon. The stem of the LIR is in length of 60 bp and the loop is formed by 10 nt (Fig. 1). There are two mismatches in the stem part and thus the matching rate of the stem is 96.7%. Aside from POTE22 genes, other POTE genes from the group 3 also contain the LIR at the same site. We therefore found eight POTE LIRs in the human genome, one on chromosome 22, two on chromosome 14 and five on chromosome 2. Intriguingly, the human genome also contains half-sized POTE LIRs that lack one arm of inverted repeat. We identified three such LIRs on human chromosome 7 and five within POTE genes belonging to the groups 1 and 2 (Fig. 2). The result of multiple alignment demonstrates that the half-sized LIRs retain one arm of about 46 nt, and a small inverted repeat of 22 nt is also missing (Figs. 1 and 2). For the half-sized LIRs on the chromosome 7, the major difference is an insertion of 14 bp in the arm and they are not within introns of POTE genes. Moreover, the regions have not been assigned with any genes at present version of human genome annotation.

In chimpanzee and rhesus monkey genomes, we identified nine and four homologous sequences respectively. The full structure of the LIR was observed in two chimpanzee chromosomes, but none in the

* Corresponding author. Fax: +852 28574672.
E-mail address: fcleung@hkucc.hku.hk (F.C.C. Leung).

0888-7543/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2009.05.005

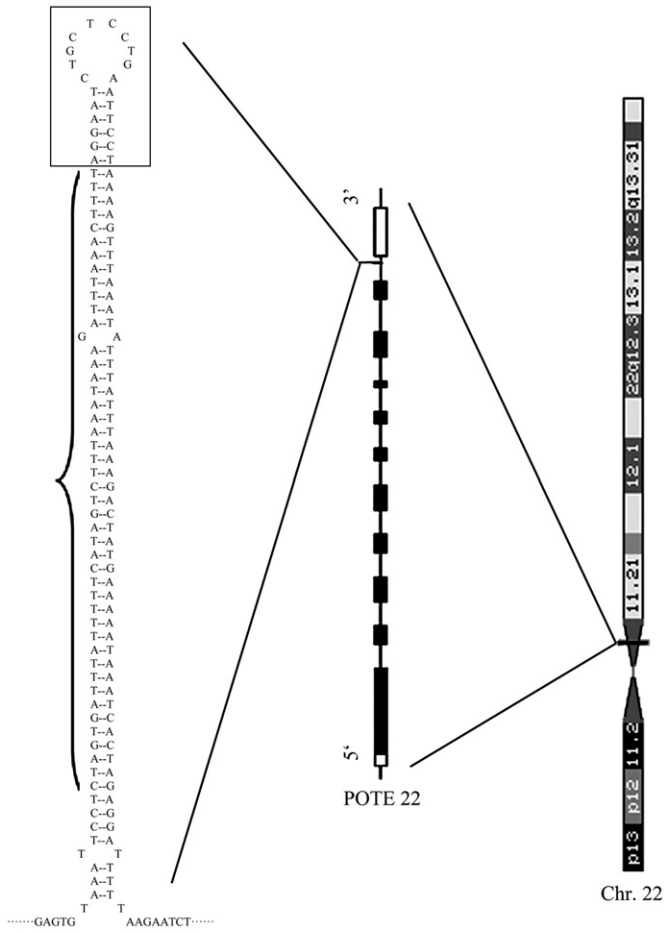

Fig. 1. LIR in POTE 22 on the human chromosome 22. Legend: The POTE22 gene was located at centromere of the human chromosome 22, and an LIR was found in the last intron of the gene. The half part of the stem in a bracket and the small stem-loop in a frame are missing in the genes with a half-sized LIR.

monkey chromosomes. They were basically located in POTE-like genes and ANKRD26-like genes (Table 1). The chimpanzee POTE genes containing the two full-sized LIRs were highly similar to POTE2α, belonging to the group 3 as well. It seems that the full-sized LIRs are restrictedly located in the group 3 of POTE family. Therefore, the difference in the number of the LIRs between chimpanzee and human is ascribed to the size of the group 3.

To confirm the difference in the LIR structure, we studied the sequence conservation at the flanking positions of the LIRs. In total, the extension of the sequence alignments to upstream and downstream was more than 500 bp, and the similarity between the sequences was between 83% and 100%. The average of the similarities obtained from pairwise comparisons was 91% with a standard deviation of 3.7%.

The homologous sequences from the three primate species were used to reconstruct a phylogenetic tree using maximum likelihood algorithm. The result shows that the sequences on orthologous regions between the three primate genomes tend to be located in the same clade (Fig. 3). Four major clades could be visualized on the phylogenetic tree, including those containing sequences on 1) chromosome 8; 2) chromosome 7; 3) chromosomes 15, 18 and 21; and 4) chromosomes 2, 14, and 22. The exceptional cases were the sequences from rhesus monkey chromosomes 3, 8 (located near 44.5 Mb), and 10 (Fig. 3). The classification of the sequences on the basis of the phylogenetic relationship is in accord with the grouping of the human POTE family if the sequences on the chromosome 7 could constitute a potential group.

Discussion

In this study, we show that the evolution of POTE gene family is accompanied with sequence variations in introns. The group 3 of the POTE family, which encodes POTE–actin fusion proteins and is believed to be derived from groups 1 and 2 [4,5], possesses an LIR in the last intron. The LIR is able to form a large stem-loop and perhaps plays an important role in expression variation and/or the evolution of the group 3.

The LIR might be functional in the following processes. Reports show that inverted repeats adjacent to exons may affect alternative splicing of exons [7–9] or knock-out efficiency of introns [10,11]. Considering the short distance to the last exon, the LIRs in POTE genes are probably important in inclusion of the last exon in mature mRNA, or have a certain linkage with the chimeric POTE–actin fusion. Moreover, the LIR might be an intronic stem-loop prepared for Dicer to cut into miRNA. According to recent progress in RNA interference, miRNAs were often identified in introns [7–9]. The spliced-out introns from an mRNA can be recruited by Dicer and processed into 20–22 nt miRNAs [12,13]. At present, we do not know if the LIR is used to produce miRNA for gene silencing. However, the possibility cannot be precluded.

Inverted repeats are motifs capable of inducing genome instability such as gene amplification, recombination and fragment reshuffling [14–17]. Palindrome structures are widespread in cancer cells, providing platform for gene amplification [18,19]. The LIR is able to further mediate amplification of POTE genes, probably accounting for over-expression of the POTE genes in cancer tissues [3]. Moreover, the POTE intronic LIR can potentially conduct duplication of exons because LIRs are motifs capable of inducing fragmental duplication [14–16,20]. Experimental works are required for understanding of the role of the intronic LIR.

Multiple sequence alignment in the present report indicates that the LIR was developed by inversely duplicating one arm of the LIR on chromosome 2, 14 or 22. In current assembly of rhesus monkey genome, we could not locate the full-sized LIR. This suggests that it was acquired by the common ancestor of chimpanzee and human, corresponding to the emerging of the POTE genes of group 3 [4,5].

In this study, the phylogenetic relationship on the basis of the intronic regions is consistent with the conclusion in previous reports using the exons of POTE genes [4,5]. The ancestral POTE gene was considered to be located on chromosome 8 [4], which is in agreement with our hypothesis for the origin of the LIR. Moreover, closer relationship between the sequences on the same chromosomes was shown in our phylogenetic tree. This suggests that the distribution of the POTE genes except those in the group 3 had been fixed in the common ancestor of rhesus monkey, chimpanzee and human. Our results showed that the spreading of the full-sized LIR on chromosomes 2, 14 and 22 was confined in chimpanzees and humans. Therefore, the first LIR should be formed in one of the three chromosomes. Interestingly, five POTE genes on human chromosome 2 are concentrated in a small pericentromeric region, implying a rapid spread on this chromosome. The subtelomeric regions of the human chromosome 2 is a well-known recombination hotspot and, as a result, contains significantly more duplicated copies compared to chimpanzee counterpart [21]. Thus, the spreading of the POTE genes in this region is probably by means of recombination. In view of evolution, this is perhaps a process of positive selection of duplication of the POTE gene associated with a LIR, and indicates a specific, functional role of the POTE genes in chimpanzees and humans. Like the POTE genes, the ancestral half-sized structure is probably primate-specific as well, because we could not find it in other mammals.

LIRs in genomes are generally derived from stretches of simple repeats, reversely-arranged repetitive elements, microRNA genes, or singlet retrotransposons with terminal inverted repeats. For those in very low frequency in a genome, we can hardly clarify how they are