Gene 425 (2008) 44–47

Contents lists available at ScienceDirect

Gene

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

Enhanced heterogeneity of the LR2 segment in the human ribosomal intergenic spacer

N.S. Kupriyanova *, D.V. Shibalev, A.S. Voronov, A.P. Ryskov

Institute of Gene Biology, Russian Academy of Sciences, 34/5, Vavilov Street, Moscow, 119334, Russia

ARTICLE INFO

Article history:
Received 4 July 2008
Received in revised form 6 August 2008
Accepted 7 August 2008
Available online 16 August 2008

Received by: A. Rynditch

Keywords:
Ribosomal DNA
Microsatellite diversity
Genetic recombination
Gene conversion

ABSTRACT

Human ribosomal intergenic spacer (rIGS) contains in its central part two highly homologous 2 kb repeats, LR1 and LR2. In this paper, we investigate heterogeneity of the variable LR2 segment ($LR2_{var}$) in the human rIGS. More than 500 $LR2_{var}$ copies from ten unrelated human genomes have been cloned and sequenced. Prolonged $(G)_n$ $(AG)_m$ compound microsatellite clusters with '*n*' and '*m*' notions fluctuating in random manner span central parts of almost all $LR2_{var}$ variants. Nucleotide sequences flanking the central microsatellite clusters are represented by more than 30 structural groups, with the two major (A and B) and six minor (C–H) ones. The analysis of sequencing data let us propose that the $LR2_{var}$ variability can be derived by various ways, including microsatellite DNA slip-strand mispairing during replication, non-equal crossover and segmental DNA exchange between $LR1_{var}$ and $LR2_{var}$ through the mechanism of gene conversion.

1. Introduction

Genomes of higher eukaryotes contain a wide variety of tandem DNA repeats. Clusters of simple sequences, microsatellites and minisatellites show considerably higher rate of evolutionary changes on a comparison with unique DNA sequences (Brohede et al., 2004; Ellegren, 2000; Ellegren, 2004; Vowles and Amos, 2004; Cox and Mirkin, 1997; Dover, 2000; Borstnic and Pumpernik, 2004; Netchvolodov et al., 2006). In the human genome, the estimated rate of point mutations is approximately 10^{-9} mutations/nucleotide/year, whilst the slippage probability is about 10^{-3} per repeat per generation. The basis for microsatellite evolution is elongations and shortenings of repeats, combined with point mutations. A recombination-based mutation process in minisatellite clusters includes duplications, deletions and gene conversion-like transfers of repeat blocks between alleles (Buard et al., 2000; Jeffreys et al., 2000; Jeffreys and May, 2004). Double-strand breaks (DSBs) initiated by staggered nicks in the tandem array with subsequent single-strand invasion of the allelic partner could account for the diversity and complexity of minisatellite rearrangements (Buard and Vergnaud, 1994).

Each ribosomal DNA (rDNA) unit consists of the transcribed region and ribosomal intergenic spacer (rIGS). The multi-copy and multiple cluster arrangement of the ribosomal genes makes the evolution of these gene systems very complicated involving different mechanisms of concerted evolution (Gonzalez and Sylvester, 2001). The human rIGS contains in its central part two ~2 kb repeats, LR1 and LR2, harboring four short hypervariable segments enriched in microsatellite clusters (Gonzalez and Sylvester, 1995). Here, we study heterogeneity of the 547 copies of the $LR2_{var}$ segment isolated from 10 human genomes. It is proposed that mechanisms providing the differences are similar to those suggested for minisatellite variability and can involve a recombination and conversional transfer between alleles.

2. Materials and methods

Total genomic DNA was isolated from blood samples using standard phenol-chlorophorm extraction with proteinase K. PCR amplification was carried out with Taq polymerase (Fermentas) on a thermal cycler Techne PHC-3 with the primers

F: (5′-TTTACCTATGTCACAGTCTTGCT-3′)
R: (5′-GAAACCCCCCTGACTCAGGTCAAG-3′)

corresponding to the rDNA nucleotide positions 23322–23344 and 23500–23523 (GenBank, U13369). PCR products were separated on 1.5% agarose gel, ligated into the pGEM-T Easy Vector (Promega) and cloned into XL1 cells. Plasmid DNA was isolated according to a standard protocol. Plasmid insertions were sequenced using fmol DNA Seq.System (Promega), with the same primers. The sequences were

Abbreviations: PCR, polymerase chain reaction; bp, base pairs; dNTP, deoxyribonucleotide triphosphates; ds, double stranded; nt, nucleotide(s); oligos, oligodeoxyribonucleotides; rDNA, DNA coding for rRNA; rRNA, ribosomal RNA; SDS, sodium dodecyl sulphate; ss, single stranded; SSC, 0.15 NaCl/0.015 M Na_3-citrate pH 7.8; EtdBr, ethidium bromide.

* Corresponding author. Fax: +7 499 1354105.
E-mail address: cuprum242@rambler.ru (N.S. Kupriyanova).

0378-1119/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2008.08.009

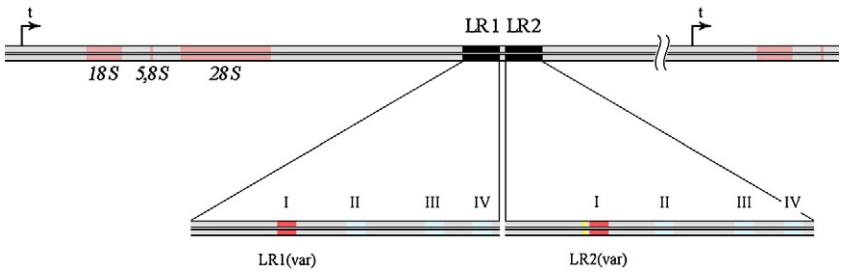

Fig. 1. A scheme of the human ribosomal DNA repeats. 18S, 28S and 5.8S rDNA regions are shown in dark gray. The curved arrows with 't' letters denote the transcription start points. The expanded region corresponds to the LR1–LR2 repeats (black rectangles). The variable LR regions (I–IV) are set off by more light colours. The regions of interest are denoted as $LR1_{var}$ and $LR2_{var}$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

aligned according to the GeneBee Program and checked manually, if it was necessary.

3. Results

3.1. Heterogeneity of the $LR2_{var}$ central $(G)_n$ $(AG)_m$ clusters

Human ribosomal intergenic spacer (rIGS) represents considerable (30 out of 43 kb) part of each of 500 rDNA monomers and contains in its central part two ~2 kb repeats, LR1 and LR2 (Gonzalez and Sylvester, 2001). Alignment of the LR1 and LR2 on the background of 88.8% similarity reveals four variable regions (Fig. 1). In this study, 547 $LR2_{var}$ DNA segments from ten unrelated human genomes with coordinates 22763–23523 apart from the rRNA transcription start point have been cloned and sequenced. Comparison of these sequences revealed their high heterogeneity and some internal structural peculiarities. Prolonged $(G)_n$ $(AG)_m$ clusters span the central parts of the $LR2s_{var}$. The '*n*' and '*m*' values vary among individual copies of the $LR2s_{var}$ in intervals 4–17, and 13–30, correspondingly, with random combinations of the $(G)_n$ and $(AG)_m$ variants. The monomer units' numbers $(G)_{8-11}$ and $(AG)_{18-20}$ are the most abundant in the total representation (Fig. 2).

3.2. Heterogeneity of $(G)_n$ $(AG)_m$ clusters' flanking regions

To systematize highly variable $LR2_{var}$ sequences, flanking the central $(G)_n$ $(AG)_m$ cluster, they have been compared, taking no account of '*n*' and '*m*' numbers, and as the result, the 30 groups have been formed (Fig. 3). One can see that the majority of $LR2_{var}$ sequences enter into groups A and B. The nucleotide sequences of the most abundant group A (82% of all the $LR2s_{var}$) are practically identical to the GenBank U13369 sequence (A1 variant), with minor variations in ~1/3 of the sequences belonging to this group (Fig. 3).

The B group (13% of all the $LR2s_{var}$) exhibits heterogeneity upstream and downstream of the central $(G)_n$ $(AG)_m$ cluster. The B1–B14 sequences reveal specific features differing them from the A group. The 5′-flanks of B type sequences contain variable $(AG)_{6-10}$ instead of conservative $(AG)_7$ arrays typical for the A group, one C–>T substitution in the fixed −18 position and short deletions upstream of the $(G)_n$ $(AG)_m$ cluster. The 3′-flanking region reveals deletions downstream of the central $(G)_n$ $(AG)_m$ cluster and base substitutions in the fixed +18, +28, +29 and +43 positions.

The C group (3% of all the $LR2s_{var}$) displays similarity to the B group upstream, and is practically identical to the A group downstream of the central $(G)_n$ $(AG)_m$ cluster that possibly reflects exchanges between A and B variants. The four variants, D, E, F, and G represented by one–two copies are irrelevant to the A–C groups.

The five uncommon sequences forming group H (H1–H5) are depleted of the central $(G)_n$ $(AG)_m$ cluster, whilst the upstream $(AG)_{6-10}$ cluster characteristic for the alleles B, is extended up to $(AG)_{18-32}$, and often contains G–>C substitutions. The 3′-part of the H1–H5 sequences harbors base substitutions, deletions, and insertions. A comparison of the H1–H5 sequences with their counterparts from the $LR1_{var}$ (GenBank, U13369) reveals identical nucleotide substitutions (Fig. 4). The reason is possibly that the two repeats, LR1 and LR2, can exchange by their DNA segments.

4. Discussion

In studies of human's minisatellite variability a convincing body of evidence have been accumulated, which suggest that along with equal exchanging of DNA sequences at the cross over points, there is an unequal conversion of one allele by the other (Buard et al., 2000). In

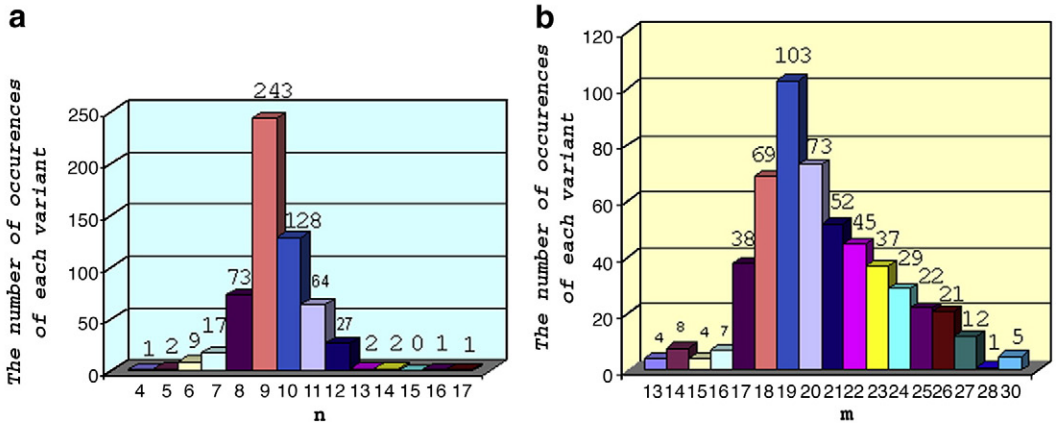

Fig. 2. A frequency of $(G)_n$ and $(AG)_m$ components of the central compound microsatellite cluster with different monomer units. a — $(G)_n$ clusters; b — $(AG)_m$ clusters.