Available online at www.sciencedirect.com
ScienceDirect
Genomics 89 (2007) 480–489
GENOMICS
www.elsevier.com/locate/ygeno

Mutation mechanisms that underlie turnover of a human telomere-adjacent segmental duplication containing an unstable minisatellite

Mark Hills [1], Jennie N. Jeyapalan, Jennifer L. Foxon, Nicola J. Royle *

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Received 31 July 2006; accepted 30 December 2006
Available online 31 January 2007

Abstract

Subterminal regions, juxtaposed to telomeres on human chromosomes, contain a high density of segmental duplications, but relatively little is known about the evolutionary processes that underlie sequence turnover in these regions. We have characterized a segmental duplication adjacent to the Xp/Yp telomere, each copy containing a hypervariable array of the DXYS14 minisatellite. Both DXYS14 repeat arrays mutate at a high rate (0.3 and 0.2% per gamete) but linkage disequilibrium analysis across 27 SNPs and a direct crossover assay show that recombination during meiosis is suppressed. Therefore instability at DXYS14a and b is dominated by intra-allelic processes or possibly conversion limited to the repeat arrays. Furthermore some chromosomes (14%) carry only one copy of the duplicon, including one DXYS14 repeat array that is also highly mutable (1.2% per gamete). To explain these and other observations, we propose there is another low-rate mutation process that causes copy number change in part or all of the duplicon.

Keywords: Telomere; Subterminal; Minisatellite; VNTR; Segmental duplication; Recombination

The distribution of homologous crossover events across human autosomes is not random during meiosis but clustered into hot spots 1–2 kb in length that are separated by regions that show strong linkage disequilibrium (LD) [1–3]. The pseudoautosomal region (PAR1) on the short arm of the sex chromosomes (Xp/Yp) spans 2.7 Mb of DNA from a Y-specific *Alu* element to the Xp/Yp telomere. This region is atypical in that it supports an obligatory crossover in male meiosis and therefore the overall male recombination rate in PAR1 is 20 times greater than the genome average [4–6]. However, LD and recombination analysis around the SHOX gene (~500 kb from the Xp/Yp telomere) has shown that most crossovers in this region of PAR1 are clustered in a recombination hot spot [7]. The 1-kb sequence immediately adjacent to the Xp/Yp telomere contains a high density of single nucleotide polymorphisms (SNPs; 1 per 65 bp among individuals from northern and western Europe) and insertion/deletion polymorphisms that show almost complete LD such that only three common haplotypes (A, B, and C) are found in this population [8]. The strong LD suggests that this region of the genome, immediately adjacent to the Xp/Yp telomere, is inert for meiotic recombination. Similar properties of high SNP density and strong LD have been found adjacent to the 12q telomere [9], suggesting this may be a general feature of telomere-adjacent regions.

The Xp/Yp and other autosomal telomeres contain many sequence-variant telomere repeat units at the beginning of the repeat array. The distribution of the variant repeats is hypervariable between alleles [8–10], indicating that telomeres have a high germ-line mutation rate, measured at 0.6% per gamete [11]. Telomeres with related variant-repeat distributions define lineages that are in strong linkage disequilibrium with SNPs in the telomere-adjacent sequence and this indicates that variability within the telomere repeat array arises from intra-allelic processes such as sister-chromatid exchange and replication slippage. The more internal subtelomeric domains of most human chromosomes contain complex patchworks of DNA sequences that are shared between nonhomologous chromosomes (interchromosomal segmental duplications) [12,13]. The duplicated segments vary considerably in length

* Corresponding author. Fax: +44 116 2523378.
E-mail address: njr@le.ac.uk (N.J. Royle).
[1] Current address: Terry Fox Laboratory, 675 West 10th Avenue, Vancouver, BC, Canada V5Z 1L3.

0888-7543/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.12.011

and in their composition of dispersed repeats, unique sequences, and genes. This subtelomeric sequence organization must have arisen recently through interchromosomal exchanges (ectopic recombination), including translocations that have likely relocated subterminal sequences and their associated telomeres [10,14–16].

The dynamics of subterminal regions up to and including the telomere are complex but it is possible that there is a barrier to homologous recombination close to the start of human telomeres. To explore this we have extended our analysis of the Xp/Yp telomere-adjacent sequence to incorporate the DXYS14 locus [17,18]. DXYS14 comprises a hypervariable minisatellite (HSVNTR, detected by the probe 29c1 [19]) based on a 31-bp GC-rich repeat unit and with an estimated mutation rate of 1.3% per gamete. Sequences homologous to the DXYS14 minisatellite are present in the chimpanzee, gorilla, and orangutan genomes, though their locations with respect to the telomere are unknown [20].

Here we present evidence that the sequence adjacent to the Xp/Yp telomere includes a small segmental duplication, each copy containing a hypervariable DXYS14 repeat array. There is some evidence that one copy of this duplicon is part of a larger segmental duplication whose sequence organization has not been fully determined (DC1321 in the TCAG genomic duplications database). Therefore our analysis is confined to the sequence in the terminal 9 kb at XpYp where the DXYS14-containing duplicons are separated by 0.8 kb of sequence rich in dispersed repeats. Both copies of the DXYS14 repeat arrays mutate in the male germ-line and yet they are embedded in a region of strong LD that extends from the telomere across both copies of the duplication. However, 14% of chromosomes carry only a single copy of the entire duplication unit, containing one minisatellite repeat array (DXYS14c) that also has a high mutation rate. We present data that indicate that a few chromosomes show additional copy number variation and carry three DXYS14 repeat arrays. The implications of these findings for mutation mechanisms within this Xp/Yp subterminal region are discussed.

Results

To extend the analysis of the region adjacent to the Xp/Yp telomere, sequence information was taken from a clone containing the Xp/Yp telomere-adjacent DNA [8,21], from a half-YAC clone (M57751 [22]), from a cosmid clone (29c1) containing part of the DXYS14 locus (X17009 [19]), and from sequence available in the GenBank and EMBL databases as part of the human genome project (Accession No. BX640545).

The Xp/Yp telomere-adjacent sequence is annotated such that the first base internal to the telomere repeat array is −1 (chrXY: 39 in Human Genome Sequence (HGS) March 2006 assembly). Dot-plot analysis of ~10 kb of sequence immediately adjacent to the Xp/Yp telomere confirmed preliminary polymerase chain reaction (PCR) results (using primers 29c1B and Tsk8W, Fig. 1a), indicating the presence of a sequence duplication (data not shown). The copy of the duplication closest to the telomere (A) extends from the −2238 position (chrXY: 2277 HGS March 2006) and comprises 1863 bp. The second copy of the duplication (B) begins at position −4954 (chrXY: 5667 HGS March 2006) from the telomere and spans 2871 bp. Both copies of the duplication contain sequences probably not found elsewhere in the genome, sequences that show partial homology to a family of human endogenous retroviruses, *Alu* sequences, and a small array of telomere-like repeats. The duplicated units differ by three insertion/deletions that result in the internal copy (B) containing an additional 1008 bp of mainly unique DNA sequence. In addition both copies of the duplication contain a variable-length array of 31 bp GC-rich repeats that comprise the DXYS14 minisatellite (from here on referred to as DXYS14a and DXYS14b in copy A and B of the duplication, respectively). The two copies of the duplication are separated by 839 bp of sequence that includes two partial *Alu* elements and part of an LTR (Fig. 1a). In addition to the insertion/deletion differences between the two copies of the duplication they show a 4% sequence divergence, indicating the duplication is ancient. Furthermore, PCR analysis of DNAs from 80 individuals (parents in the Centre d'Etude du Polymorphism Humain (CEPH) family DNA panel) with the 29c1B and Tsk8W primers, which generate amplicons of different lengths from the two copies of the duplication, showed that some chromosomes carry only a single copy of the duplication unit (see below).

SNP detection adjacent to the Xp/Yp telomere

A high density of DNA sequence polymorphisms in the terminal 1.2 kb adjacent to the Xp/Yp telomere has been reported previously [8,9]. Sequence analysis was undertaken across the extended telomere-adjacent region to determine the density and distribution of DNA sequence polymorphisms. Seventeen candidate SNPs were identified between the −1890 SNP and the DXYS14b repeat array, some of which are present in dbSNP, and one novel SNP was identified internal to DXYS14b at the −7152 position (Fig. 1a).

Linkage disequilibrium across the region

The polymorphic sites in the Xp/Yp telomere-adjacent sequence, from the −1890 SNP (previously reported as −1888) to the −13 SNP next to the telomere, are in almost complete LD, suggesting there has been very little recombination in this region during recent evolution [8]. To determine whether strong LD extends into the region encompassing the newly identified SNPs that span the DXYS14a and DXYS14b repeat arrays, allele-specific oligonucleotide (ASO) hybridization assays were developed for the new SNPs. Some of the DNA samples generated more than one product with the 29c1M and 29c1X primers or a product of unexpected, larger size. The products of expected size contained the sequences described here but sequence analysis of the larger products did not resolve the origin of these fragments. To avoid any incorrect genotyping, the four candidate SNPs in this region (−3640, −3644, −4781, and −4838) were excluded from subsequent