

ARTICLE

Human pseudoautosomal boundary-like sequences: expression and involvement in evolutionary formation of the present-day pseudoautosomal boundary of human sex chromosomes

Tatsuo Fukagawa, Yasukazu Nakamura, Katsuzumi Okumura¹, Masahiro Nogami¹, Asako Ando², Hidetoshi Inoko², Naruya Saitou and Toshimichi Ikemura*

Department of Evolutionary Genetics, National Institute of Genetics, and The Graduate University for Advanced Studies, Mishima, Shizuoka-ken 411, Japan, ¹Faculty of Bioresources, Mie University, Tsu, Mie-ken 514, Japan and ²School of Medicine, Tokai University, Bohseidai, Isehara, Kanagawa-ken 259-11, Japan

Received August 23, 1995; Revised and Accepted October 10, 1995

The human genome is composed of long-range mosaic structures of G+C% (GC%), which are thought to be related to chromosome bands. We previously identified a boundary of Mb-level domains of GC% mosaic structures in the human major histocompatibility complex (MHC) and found in the domain boundary a sequence very similar to pseudoautosomal boundary (PAB) sequences of human sex chromosomes. We designated it 'PABL' and found many PABLs in the human genome. By analysis of six genomic and six transcribed PABLs, a core and consensus sequence of about 650 nt was defined; the 3'- and 5'-edges of the PABLs were strictly conserved. Northern blot analysis showed sizes of PABL transcripts to be 5–10 kb in length. Divergence time of PABLs was estimated to be 60–120 million years ago by analysis of human PABLs and PABXY1 of seven primates, and the evolutionary rates deduced showed PABLs to have been under selective constraints. A model for evolutionary formation of the present pseudoautosomal boundary was proposed by postulation of illegitimate recombination between two PABLs.

INTRODUCTION

The human genome, like the genomes of warm-blooded vertebrates in general, has long-range G+C% (GC%) mosaic structures related to chromosome bands: Giemsa-dark G bands are composed mainly of AT-rich sequences, T bands (a subgroup of Giemsa-pale R bands) are of GC-rich sequences, and ordinary R bands are heterogeneous and appear to be intermediate (1–6). Bernardi *et al.* called the GC% mosaic domains 'isochores' (1). DNA replication timing, gene density, CpG island density, codon usage, chromosome condensation, repeat sequence density, and chromosome behavior such as recombination are related to chromosome bands and to long-range GC% mosaic domains (5–12). Because chromosome bands are structures observed by microscopes, precise location of their boundaries may seem meaningless. However, considering the various genome beha-

vors connected with chromosome bands, we may precisely locate their boundaries by using informative landmarks on genomic DNA. Boundaries may be structurally assigned where there are clear GC% transition points, and characteristic signals may be found that punctuate and/or differentiate distinct functions, such as a signal that switches from early to late DNA replication. We previously reported a boundary of long-range GC% mosaic domains in the human major histocompatibility complex (MHC) disclosing a sharp GC% transition, and found in the boundary region high sequence similarity with the pseudoautosomal boundary (PAB) sequence in the short arms of human sex chromosomes (13).

Human sex chromosomes have two functionally distinct regions, sex-specific sequences and pseudoautosomal regions (PARs). X and Y chromosomes exchange DNA sequences through homologous recombination in PARs during each male

*To whom correspondence should be addressed

Sequence data from this article have been deposited with the GenBank/EMBL/DDBJ Data Libraries under Accession Nos. D55638 (Bc4), D55639 (Mo1), D55640 (Mo2), D55641 (Sk13), D55642 (PABLSp2), D55643 (Sp2), D55644 (Sp3), D63517 (PABL consensus sequence)

meiosis, and PAR sequences in the two chromosomes are practically identical because of the obligatory recombination (14–17). There are two PARs for human sex chromosomes: PAR1 is at the distal ends of the short arms of the X and Y chromosomes and PAR2 is at the ends of the long arms (17,18). The interface between the PAR1 of 2.6 Mb and the sex-specific region is the pseudoautosomal boundary, PAB1, and therefore PAB1 is the centromeric limit to recombination in PAR1. Ellis and Goodfellow (16) and Ellis *et al.* (19,20) reported sequences around the interface, PABX1 and PABY1 sequences (abbreviated PABXY1). A sequence found in the boundary of long-range GC% mosaic domains in the MHC is very similar to PABXY1 being the functional interface in sex chromosomes. We designated the sequence found in the MHC 'pseudoautosomal boundary-like sequence 1', abbreviated PABL1 (13). With the PABL1 segment as a probe, many copies of pseudoautosomal boundary-like sequences (PABLs) were detected by Southern blot hybridization against genomic DNAs. In the present study we characterized both genomic and transcribed PABLs.

RESULTS

Core sequences of genomic PABLs

Using the PABL1 segment found in the human MHC as a probe, we isolated about 150 independent cosmid clones having PABLs. Figure 1 shows alignment of genomic PABLs sequenced thus far and PABXY1 plus their flanking sequences; the sequence orientation for PABXY1 is from telomere to centromere (i.e., from PAR1 to the sex-specific region). Similarity for all pairs of PABLs and PABXY1 were high (~80% nucleotide identity), and the 3'-terminus of the similar region among all PABLs corresponds closely to the PABXY1 homology terminus reported by Ellis *et al.* (19,20) as where the X and Y chromosome sequences diverge completely (sex-specific in Fig. 2).

In the case of the 5'-terminus of PABXY1 sequences there was no reason to assign a specific site because the upstream 2.6 Mb sequence of PAR1 should be practically identical in sex chromosomes. In fact, the 5'-terminus of PABXY1 sequences reported (19) was an *EcoRI* site chosen arbitrarily in PAR1. Fukagawa *et al.* (13) noted that, even if PABXY1 and PABLs correspond to a certain functional signal, the 5'-end of the respective signal cannot be elucidated by comparison of the X and Y sequences, and that only comparison with or among PABLs will define the 5'-terminus of the hypothesized signal. Recently Dr N. A. Ellis kindly supplied us with his unpublished PAR1 sequence upstream of the *EcoRI* site. PABXY1 in Figure 1 includes this sequence. Multiple alignments of Figure 1 showed that sequence similarity among all PABLs and PABXY1 actually extended 200 nt upstream of the *EcoRI* site of PABXY1. We propose that a total of 650 nt including the 200 nt extended is the core form of PABLs. Figure 2 is a diagram of their structure. Strict conservation of both termini for all PABLs and PABXY1 may be due to their function.

The transcripts of PABLs

Although multiple copies of PABLs were found in the human genome, genomic sequences in the recent GenBank (Release 87 including update data; 1995) exhibiting significant similarity with the PABLs were confined to PABXY1. In the case of expressed sequence tag (EST) sequences, however, eleven ESTs

were similar (~80% nucleotide identity) to separate portions of PABLs (Fig. 2), suggesting that some, if not all, PABLs are transcribable and presumably have functions. For identification of the characteristics of the predicted PABL transcripts, six human cDNA libraries of different tissues and cells were screened with the PABL1 segment as a probe. Clones were obtained from all six libraries, and inserts of all 20 cDNA clones analyzed were longer than the PABL size. We then sequenced the following six cDNA clones: Mo1 and Mo2 from the monocyte library; Bc4 from the B-cell library; Sk13 from the skin library; and Sp2 and Sp3 from the spleen library. Alignment around the 5'-terminal portion of all available PABLs found in the cDNA clones including ESTs (T92306 and R12279) is shown in Figure 3A. The similarity terminus of the 5'-portion of these transcribed PABLs corresponds closely to that defined for genomic PABLs. It should first be noted that a portion of the Mo2 cDNA sequence was practically identical to one of the ESTs mentioned above, HSAAA-AWAH. The identity portion spanned not only the PABL sequence but also the entire flanking region of 266 nt HSAAA-AWAH, predicting the EST to correspond to a portion of the Mo2 transcript and the transcript to be fairly abundant. Sequences of three recently reported ESTs (R12279, R07404, and T99392, derived from at least two independent cDNA libraries) were almost identical with each other, indicating the transcripts are also abundant.

Figure 3B shows alignment around the 3' portion of the transcribed PABLs and one EST (T47905). Again, the similarity terminus of these transcribed PABLs closely corresponds to that defined for genomic PABLs and thus to the homology terminus between PABX1 and PABY1. Therefore, the results of Figures 1 and 3 showed conservation of both 5'- and 3'-termini of genomic and transcribed PABLs. Open reading frames (ORFs) with significant sizes could not be found for the obtained cDNAs, for the core sequences or their flanking sequences. We searched these cDNA sequences, using the BLASTX program, against the protein sequence database compiled by the Human Genome Center of Japan. Little if any similarity with known proteins was detected. GRAIL, a computer program that predicts protein-coding ORFs in human DNA (21), did not detect reliable protein-coding capacity. For identification of the genomic structures of the transcribed PABLs, a λ -phage library of human genomic DNA was screened with a probe of the Sp2 cDNA fragment deprived of the PABL core. The genomic PABLSp2 sequence shown in Figure 1 is derived from the phage clone thus isolated. The sequences of the 5'-flank of about 50 nt and the 3'-flank of about 300 nt so far known for Sp2 cDNA were identical to those of the genomic PABLSp2 clone, showing that there were no intron/exon structures around this PABL core.

Northern hybridization analysis

The six cDNAs of PABLs analyzed were longer than the PABL size. For estimation of the intact sizes of PABL transcripts, northern blot analysis of human total or polyA⁺ RNA fraction was done with the PABL1 probe. Figure 4A shows the results for the total RNA extracted from B-cell line GM01416D. Broad bands mainly with mobility slower than that of 28S rRNA and estimated to be 5–10 kb in length were detected. The results for the polyA⁺ RNA fractions were practically the same for total RNA (Fig. 4B). All samples shown in Figure 4 gave signals, although expression levels seemed to differ. The broad bands shown are likely to

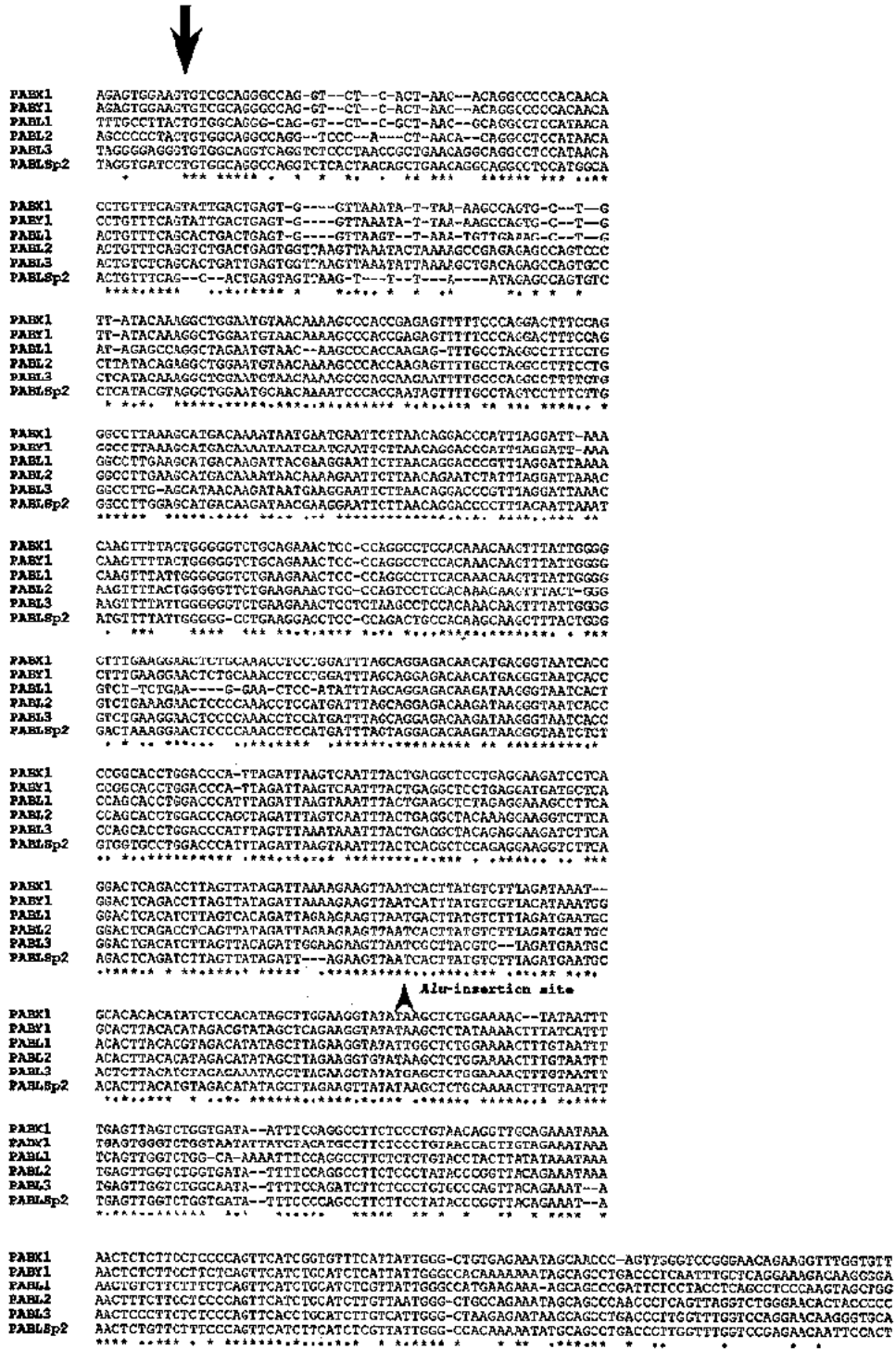


Figure 1. Comparison of genomic sequences of PABXY1 and PABLs. Multiple alignment of nucleotide sequences was performed with the MALIGN program. An Alu sequence present in PABY1 was omitted in the alignment. Identical nucleotides are marked with asterisks, and positions where five or six nucleotides are identical are marked with dots. PABXY1 sequences upstream of the EcoRI site were supplied by Dr Nathan A. Ellis. Arrows show termini of the proposed core unit of PABLs.

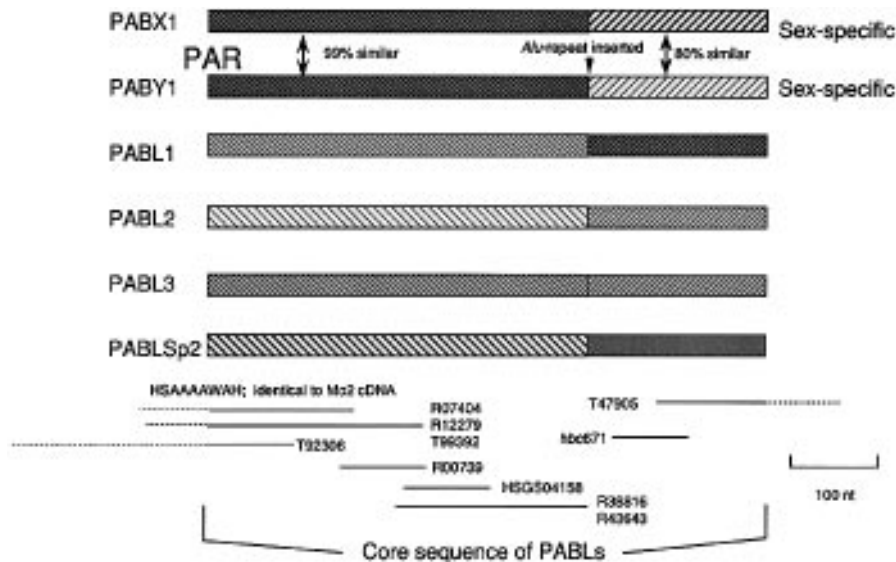


Figure 2. Genomic structure of PABXY1 and PABLs. For both PABXY1 and PABLs, regions before and after the *Alu*-insertion site are marked separately. The core unit of PABL (~650 nt) is present in all sequences, and their flanking sequences are divergent. In the case of PABXY1, the 3' flank constitutes the sex specific region. EST sequences showing evident similarity with PABLs are shown by horizontal lines: GenBank loci HSAAAAWAH (accession no. Z19872), hbc671 (T11103), HSGS04158 (D25790), and GenBank loci and accession nos. R07404, R12279, T99392, T47905, T92306, R38816, and R43643.

correspond to many transcripts hybridized with PABL sequences, and their length being long was consistent with the finding that the PABL cDNAs were longer than the PABL size. For detection of a single transcript corresponding to a unique cDNA, a Sp2 cDNA fragment deprived of its PABL core was used as the hybridization probe, and in this case a single sharp band of 7.5 kb was detected (Fig. 4C).

Phylogenetic relationships among PABLs and PABXY1

Obligatory pairing and crossover in the PAR ensure accurate segregation of the sex chromosomes during male meiosis. An unusually high rate of homologous recombination in PAR1 (20-fold the genome average) seems to arise from a mechanism that promotes obligatory physical association and the successive crossover between the sex chromosomes. The strict limit to terminate the high frequency recombination in PAR1 is the pseudoautosomal boundary 1 (PAB1). An *Alu* element is situated in PAB1 of the human Y chromosome, PABY1, but not in PABX1 (19). Ellis *et al.* defined first the *Alu* element as the strict boundary of PAR1. A later study of the boundaries of PAR1 in Old World monkeys, however, found no *Alu* repeats on the PABY1 (20). In spite of the lack of the *Alu* element, they proposed the *Alu*-insertion site itself as the strict boundary of PAR1 for the following reasons. In pair-wise comparisons of the X and Y boundary sequences within each species of several hominoids and Old World monkeys, they found sequences about 220 nt long downstream of the *Alu* insertion site to be more divergent between sex chromosomes than the sequences of the *Alu*-upstream region (~78% vs. ~97% nucleotide identity; see Fig. 2 for human PABXY1), whether or not the *Alu* element was present. By extensive analysis of the nucleotide substitution patterns, they concluded that the *Alu*-insertion site, which corresponds to the abrupt transition between the high- and reduced homology

regions, is the strict limit for high-frequency recombination in PAR1 and thus the exact boundary of PAR1 (20). The position of the boundary in Old World monkeys and hominoids was practically the same. This shows that the limit of PAR1 recombination was at the *Alu*-insertion site before divergence of lineages of Old World monkeys and great apes, and that an *Alu* element was inserted into the preexisting boundary of the Y chromosome in the great ape lineage. As did Ellis *et al.*, this site is called the '*Alu*-insertion site' in this paper whether or not the *Alu* element is present; the high- and reduced-homology regions are thus called the *Alu*-upstream and *Alu*-downstream regions, respectively. Downstream of the *Alu*-downstream region, there is no similarity between X and Y sequences, which are called the sex-specific regions in Figure 2.

Sequences around the *Alu*-insertion site being the exact interface that divides sex chromosomes into two functionally distinct regions are known to be strictly conserved in species of hominoids and Old World monkeys (20). At the insertion site there is a 10 nt sequence strictly conserved even among PABLs, although the *Alu* element is absent (see Fig. 1). Strict sequence conservation around this *Alu*-insertion site, as well as strict conservation of the 5'- and 3'-termini, during evolution may be related to possible functions of PABXY1 and PABLs. In our previous paper, we reported a rather unexpected finding that the sequence similarity of the *Alu*-downstream portion of PABX1 or PABY1 with the PABL2 (82.2 or 80.5% nucleotide identity) is higher than that between PABX1 and PABY1 (77.8% identity) and the similarity among PABX1, PABY1, and PABL1 is equivalent (13). For clarifying evolutionary processes that formed the present PABXY1 and PABLs and investigating functional constraints on the sequences, phylogenetic relationships and evolutionary rates were examined for all available PABLs. The reported PABXY1 sequences of great apes and Old World monkeys (20) were included for estimation of the

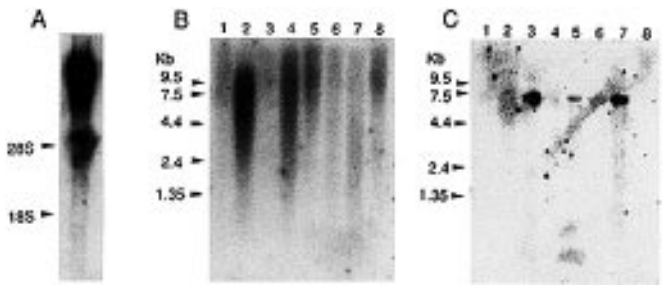
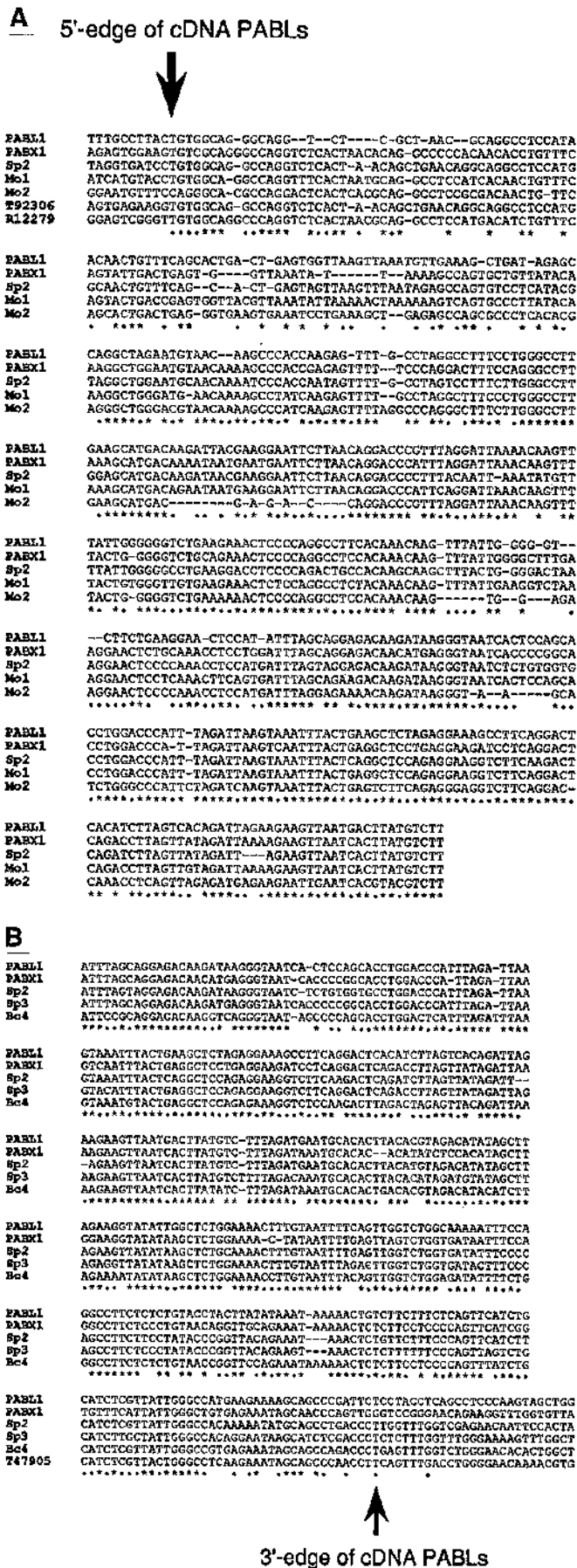


Figure 4. Northern blot analysis of human total or polyA⁺ RNA. (A) Total RNA was extracted from GM01416D cells by the AGPC method (39). Twenty-five µg of the total RNA was electrophoresed on a 1% agarose gel in a buffer containing 6% formaldehyde, 20 mM MOPS (pH 7.0), and 5 mM sodium acetate, and then transferred to a Hybond-N⁺ membrane. The probe used for hybridization was a 598 nt PCR fragment of PABL1. Hybridization was done in solution containing 5 × SSPE, 10 × Denhardt's solution, 100 µg/ml freshly denatured salmon sperm DNA, and 2% SDS. (B) Two µg of polyA⁺ RNA fractions derived from different tissues (lane 1, spleen; lane 2, thymus; lane 3, prostate gland; lane 4, testis; lane 5, ovary; lane 6, small intestine; lane 7, colon; lane 8, peripheral blood leukocytes) were tested. The RNAs were obtained from Clontech (Palo Alto, CA), and the probe was the PABL1 segment described above. (C) The polyA⁺ RNA filter in (B) was used after removal of the PABL1 probe. The probe used for this hybridization was a unique 800 nt fragment of Sp2 cDNA deprived of the PABL sequence.

evolutionary rates. Divergence between great apes and Old World monkeys is postulated to have occurred 25 million years ago. Because of the PAR characteristic, PABX1 and PABY1 sequences upstream of the *Alu*-insertion site are practically identical within a single species. For avoidance of complications arising from this peculiar characteristic, individual PABXY1 and PABLs were divided into two regions and unrooted phylogenetic trees were separately constructed by the neighbor-joining method (Fig. 5). Trees obtained from evolutionary distances on the basis of one- and two-parameter methods were identical in their branching patterns or topologies, and branch lengths were nearly the same.

The downstream portion, and thus the reduced-homology portion for sex chromosomes, was analyzed first. PABX1 and PABY1 of all species examined were separated into two distinct

Figure 3. (A) Comparison of the 5'-region of PABLs of cDNA clones with that of PABL1 and PABX1. Positions of identical nucleotides are marked with asterisks, and positions where four of five nucleotides are identical are marked with dots. Arrow shows the 5'-edge of the proposed core unit of PABLs. Near the 5'-edge, two EST sequences, T92306 and R12279, were added to the alignment. The multiple alignment analyzing only cDNA and EST sequences showed the 5'-terminus of similarity to be practically identical to that for the genomic PABLs. (B) Comparison of the 3'-region of PABLs of cDNA clones with that of PABL1 and PABX1. Near the 3'-edge, one EST, T47905, was added to the alignment. The multiple alignment analyzing only cDNA and EST sequences showed that the 3' terminus of similarity approximately corresponds to that for the genomic PABLs. One edge of cDNAs and ESTs were often within the PABL core. This may be due to the stable secondary structure predicted for PABLs which presumably inhibits cDNA extension by reverse transcriptase: the most stable secondary structure for PABLs was calculated with the GCGFOLD program managed by the University of Wisconsin Genetics Computer Group (UWCG), and energy levels found for individual PABLs were equivalent to or lower than those of RNAs known to form stable secondary structures such as mitochondrial rRNAs and 7SL RNA (data not shown).

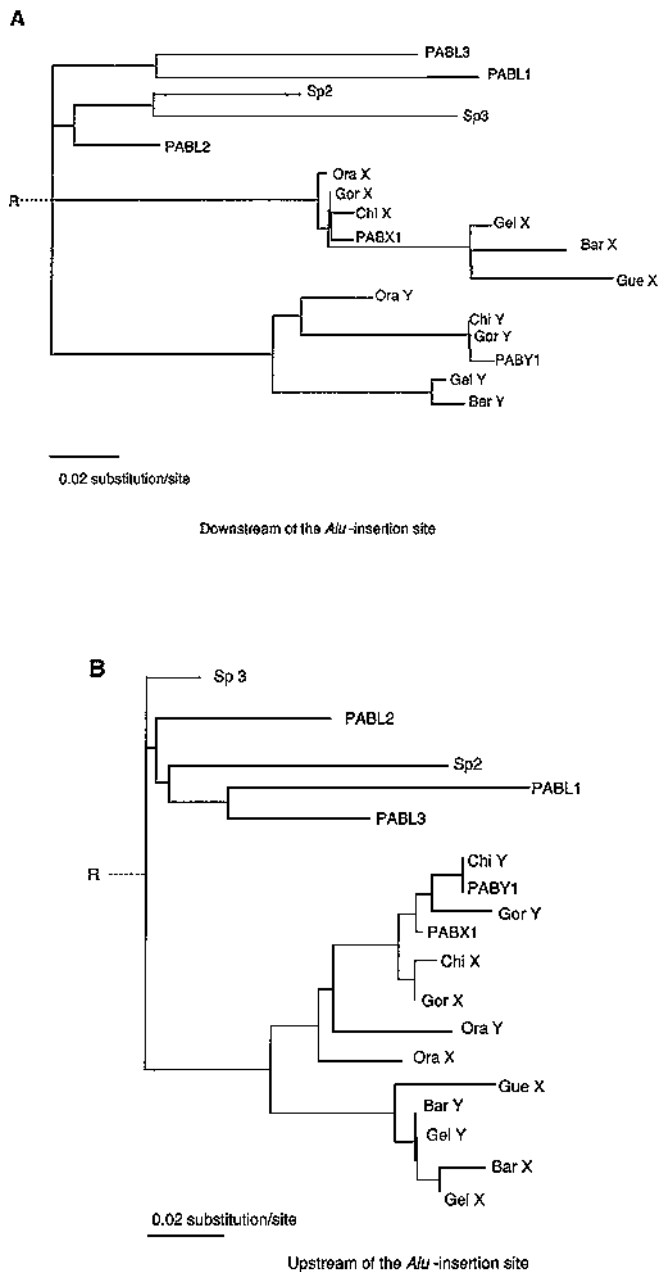


Figure 5. Phylogenetic trees of PABXY1 and PABL sequences, based on the downstream (A) and upstream (B) portions to the *Alu*-insertion site. Trees were constructed for PABXY1 of four hominoids and three Old World monkeys and five human PABLs by the neighbor-joining method. Mainly because of limited availability of PABXY1 sequences of great apes and Old World monkeys, the portions analyzed (151 nt for downstream and 177 nt for upstream) were shorter than those of the PABL core. Abbreviations for the species are as follows: *Homo sapiens* (human; Hum), *Pan troglodytes* (chimpanzee; Chi), *Gorilla gorilla* (gorilla; Gor), *Pongo pygmaeus* (orangutan; Ora), *Theropithecus gelada* (gelada baboon; Gel), *Macaca sylvanus* (barbary macaque; Bar), and *Cercopithecus cephus* (moustached guenon; Gue). The root of trees was predicted by the UPGMA method (40). Branch lengths were proportional to the estimated number of nucleotide substitutions. Evolutionary distances (number of nucleotide substitutions) were estimated by Kimura's two parameter method (41), and these distances were used to construct neighbor-joining trees. Bootstrap probabilities, based on 1000 resamplings, were calculated for each internal branch of neighbor-joining trees with the NJBOOT2 program (kindly provided by Dr K. Tamura, Tokyo Metropolitan University).

groups and the topology within each group reflects the known phylogenetic relationships of the species (Fig. 5A). The divergence between PABX1 and PABY1 sequences was similar in extent to the divergence of PABX1 (or PABY1) from PABLs. This is consistent with our previous finding obtained analyzing PABL1 and 2, as well as with the suggestion of Ellis *et al.* (20) that genetic contact resulting in sequence homogenization between sex chromosomes did not occur in the *Alu*-downstream portion after the divergence of great ape and Old World monkey lineages and the strict limit of PAR1 recombination was at the *Alu*-insertion site in the ancestral species of all extant higher primates. From the evolutionary distance between human PABX1 and Old World monkey PABX1, as well as that for PABY1, the divergence times of PABXY1 and thus of PABLs was estimated to be 60–120 million years. This estimate is consistent with the finding that PABLs are present in the bovine genome (13). Based on the divergence time, evolutionary rates of individual PABLs were estimated (Table 1).

Table 1. Evolutionary rates of PABLs and PABXY1

	Rates	
	Upstream	Downstream
PABL1	1.2 ± 0.4	1.5 ± 0.5
PABL2	0.6 ± 0.2	0.4 ± 0.1
PABL3	0.7 ± 0.3	1.3 ± 0.4
Sp2	0.9 ± 0.3	0.8 ± 0.3
Sp3	0.1 ± 0.05	1.5 ± 0.5
PABX1	0.9 ± 0.3	1.0 ± 0.4
PABY1	0.9 ± 0.3	1.5 ± 0.5

All rates are in units of 10^{-9} substitutions per site per year. Divergence time of PABLs is assumed to be 60–120 million years.

We then analyzed the upstream portion of the *Alu*-insertion site (Fig. 5B). As expected from the PAR characteristic, PABX1 and PABY1 do not separate into two groups. Evolutionary distances of several PABLs such as PABL2, PABL3, and PABLSp3 (PABL portion of Sp3 sequence) differ significantly between the upstream and downstream portions of the *Alu*-insertion site (Fig. 5A vs. B). This difference suggests that the *Alu*-insertion site corresponds to recombinational and/or functional interface within a PABL as observed for PABXY1 although the *Alu* element itself is absent. Evolutionary rates for the upstream portion estimated from the PABL divergence time obtained for the downstream portion are also listed in Table 1. The evolutionary rates of some PABLs were estimated to be far smaller than 1×10^{-9} substitutions per site per year (e.g. 0.1×10^{-9} substitutions per site per year for the *Alu*-upstream portion of PABLSp3). This finding indicates that the sequences evolved more slowly than typical noncoding regions, since the averages of evolutionary rates for mammalian pseudogenes and introns of transcribed genes have been estimated to be 4.9×10^{-9} and 3.7×10^{-9} substitutions per site per year, respectively (22). These results suggest that PABLs and PABXY1 have been under selective constraints and presumably have functions. We also analyzed a 300 nt X-specific sequence immediately downstream of PABX1 (sex-specific in Fig. 2), which has been reported in hominoids and Old World monkeys (20). Pairwise-comparison of the sequence

between species and the same comparisons for the *Alu*-upstream or -downstream PABX1 sequence were conducted for estimation of the nucleotide divergence in the regions. In most cases, nucleotide divergence in the PABX1 sequences was lower than that of the X-specific sequence, presenting evidence for evolutionary and functional constraints on the PABX1 sequence (Table 2).

Table 2. Percentage of divergence in pairwise comparison of sequences within or around pseudoautosomal boundary 1 of X chromosome

	PABL sequence		Sex-specific
	Upstream of <i>Alu</i> insertion	Downstream of <i>Alu</i> insertion	
HumX/ChiX	1.5	1.2	1.0
HumX/GorX	1.0	0.6	2.1
ChiX/GorX	0.5	0.6	1.7
HumX/OraX	4.1	1.2	5.5
ChiX/OraX	4.6	1.2	5.1
GorX/OraX	4.1	0.6	6.2
HumX/BarX	10.7	8.1	11.6
HumX/GelX	9.6	6.9	10.8
HumX/GueX	11.2	9.4	14.2
ChiX/BarX	10.2	8.1	11.2
ChiX/GelX	9.1	6.9	10.5
ChiX/GueX	9.6	9.4	13.9
GorX/BarX	10.7	7.5	12.2
GorX/GelX	9.6	6.2	11.5
GorX/GueX	10.2	8.8	14.5
OraX/BarX	9.6	8.1	10.8
OraX/GelX	8.6	6.9	10.6
OraX/GueX	10.2	9.4	12.8
BarX/GelX	2.0	4.4	3.8
BarX/GueX	4.6	6.2	5.1
GelX/GueX	3.6	5.6	5.8

Abbreviations for species are listed in Figure 5 legend.

More than ten PABL sequences (genomic PABLs, PABXY1, and transcribed PABLs) have been studied, and their sequences are about 80% identical or more. This high similarity allowed deduction of a consensus sequence of 646 nt which we deposited with GenBank/EMBL/DDBJ (accession no. D63517); a unique base could be assigned to 626 nucleotide positions in the 646 nt PABL core. Tree topology in Figure 5 indicates the consensus sequence is most likely related to their ancestor sequence. When the consensus sequence was aligned with PABLSp3, divergence found for the *Alu* downstream portion (11.9% nucleotide difference) was higher than that for the upstream portion (5.7% difference). This difference and results shown in Figure 5 and Table 1 support the model that the *Alu* insertion site is a recombinational and/or functional interface within a PABL as with PABXY1.

DISCUSSION

Boundaries of long-range GC% mosaic domains

At a standard 850 band level, the human MHC harboring PABL1 is on a wide R band, 6p21.3, which has been assigned to a T-type R band (T bands); T bands are an evidently heat-stable subgroup of R bands and mainly composed of GC-rich sequences. By high-resolution banding, a thin G positive subband, 6p21.32, was found within the MHC (23). On the basis of detailed base composition analysis, Fukagawa *et al.* (13) predicted that the genome portion harboring the evidently AT-rich 200 kb around the junction between classes II and III constitutes a portion of the thin G positive subband. PABL1 was within the AT-rich domain but close to the domain edge, suggesting PABL1 is located near a chromosome band boundary. Considering the wide range of functional behaviors of chromosome bands and GC% mosaic domains, their boundaries are most probably composed of multiple signals and structures related with multiple functions. Near the GC% mosaic boundary in the MHC the following characteristic structures have been found: on the GC-rich side a dense 20 kb *Alu* cluster and on the GC-poor side a dense 30 kb LINE-1 cluster and PABL1 (13). If these characteristics are found for other boundaries, it will add to our comprehensive understanding of both DNA sequences and chromosome structures. We focus here on GC% distribution and band structures around PABXY1. PAR1 has been assigned to an R band (Xp22.33 and Yp11.32), and judging from the high density of CpG islands a major portion of 2.6 Mb of PAR1 is thought GC-rich (17). The boundaries with the neighboring G-positive subbands, Xp22.32 and Yp11.31, appear rather close to PABXY1 (17,24) (Fig. 6). PABXY1 and their neighboring sex-specific sequences, including SRY, RPS4Y, and ZFY located downstream of PABY1, are AT-rich (20). We therefore previously predicted that PABXY1 is near a boundary of the long-range GC% mosaic domains and of chromosome bands in the sex chromosomes (13). Very recently, analyzing a contiguous 41 kb sequence harboring PABY1 (22.5 kb of PARY1 and 18.5 kb of Y-specific sequence), Whitfield *et al.* (25) showed that the PARY1 and Y-specific portions differ dramatically in GC%. It thus becomes clear that not only PABL1 in the MHC but also PABY1 are located in the GC% mosaic boundaries and therefore genome characteristics in regions harboring these elements are similar.

To further study genome features of regions with PABLs, we recently characterized a 3 Mb portion harboring PABL2 by analyzing YAC and cosmid clones in the respective regions. The 3 Mb region was found to be composed of long-range GC% mosaic domains and PABL2 was in the domain boundary region (our unpublished data). Therefore, genome characteristics around PABL2 are also analogous to those around PABL1 and PABY1. For identification of chromosome locations of other PABLs, we used a standard fluorescence *in situ* hybridization (FISH) onto metaphase chromosomes. Most of the ten PABLs mapped already were on T-type and/or terminal R bands in which there are internal narrow G subbands (our unpublished data). Characteristics of the genome portions harboring PABLs appear again to resemble each other.

Possible functions of PABLs and PABXY1

Functionally important molecules evolve more slowly than less important molecules (26). PABLs probably have biological

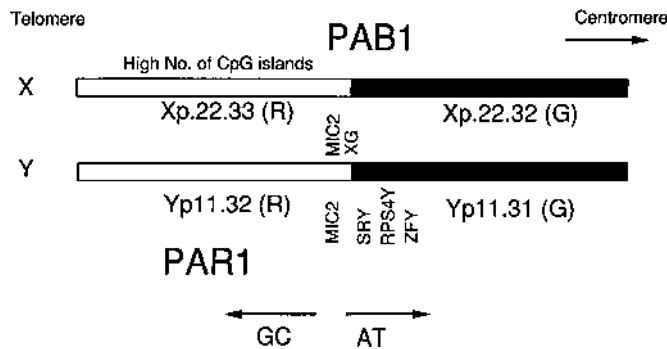


Figure 6. Chromosome bands and long-range GC% mosaic structures around the pseudoautosomal boundary of the short arms of the human sex chromosomes. Pseudoautosomal regions (PAR1) have been assigned to R bands (Xp22.33 and Yp11.32) and, judging from the high density of CpG islands, a major portion of PAR1 is thought to be GC-rich. The boundaries with the neighboring Giemsa-positive subbands, Xp22.32 and Yp11.31, appear rather close to PABXY1 and tentatively placed near PAB1 according to Mandel *et al.* (24). PABXY1 and their neighboring sex-specific sequences, including SRY, RPS4Y, and ZFY are known to be AT-rich. These observations suggest the PABXY1 to be in a boundary of the long-range GC% mosaic domains.

functions because they are transcribable and their evolutionary rates were slower than nonfunctional regions. Concerning the functions of boundaries of chromosome bands and GC%-mosaic domains, the switching of DNA replication timing is an important candidate (11,13). Multi-color FISH of interphase nuclei (27), in which contiguous λ phage and cosmid clones in the MHC (13,28) were used as probes, showed that PAB1 is located in a switching region of DNA replication timing (Okumura *et al.*, in preparation). Major portions of the methylated and transcriptionally inactive X chromosome replicate very late in S phase, but PAR1 escaping from the X-inactivation is thought to replicate much earlier (29,30). Therefore, a switch in DNA-replication timing, at least in the inactivated X chromosome, presumably occurs near PAB1. PABLs including PABXY1 may be related to possible signals for DNA replication timing such as a pause signal for the replication. Considering that high frequency-recombination in PAR1 between sex chromosomes terminates abruptly at PAB1, PABLs are expected to be involved in a certain process of recombination. A model for evolutionary formation of the present pseudoautosomal boundaries is proposed below by postulation of illegitimate recombination between two PABLs.

Characteristics of PABL transcripts

PABX1 was found within an intron of PBDX (pseudoautosomal boundary divided on the X chromosome), which encodes the Xg^a antigen and was recently renamed XG (31,32). When RT-PCR analyses were done on human total RNA from a B-cell line, primers designed for the present cDNAs reproducibly gave PCR products with the expected sizes under standard conditions (data not shown). However, primers designed for the PABX1 region showed only occasionally a faint band of the expected size under the same or modified conditions. RNA molecules corresponding to the present cDNAs may be more abundant or have longer life-span than the RNA of the XG intron.

Significant ORFs were not found for the present cDNA sequences, for either PABLs or their flanks. Computer searches could assign no protein-coding capacity. It should also be noted

here that intact size of PABL transcripts ranges from 5 to 10 kb but the length of cDNA sequences obtained in this study is 2 kb or less. Therefore we can not exclude the possibility that protein-coding regions are present in the unanalyzed portions. Alternatively, these results may suggest that the functional form of the transcripts is RNA molecules. PABL1 and PABXY1 were both found in experiments that searched for molecular structures and signals differentiating global characteristics in the human genome. Large PABL transcripts may be necessary for recognition or differentiation of global genome characteristics. XIST RNA which is 17 kb long has been thought important in such determination and differentiation in the inactivated X chromosome (33).

Evolutionary processes in formation of sex-chromosome PABLs

Recently Ellis *et al.* (31) proposed a model for the evolutionary formation of the present-day pseudoautosomal boundary of the short arms, PABXY1. They hypothesized a pericentric inversion of the Y chromosome with one break point in ancestral XG and the other break point 5 kb distal to the ancestral SRY believed to be on the earlier long arm (see legend of Fig. 7). Here we propose a model in which the hypothesized inversion occurred by illegitimate recombination between the two PABL elements, one PABL in the ancestral XG and the other near the ancestral SRY; before recombination, the two elements on the earlier Y chromosome were ordinary PABLs (Fig. 7A). After recombination, the characteristics of a pseudoautosomal boundary was acquired and the recombinant PABL became PABY1 (Fig. 7C).

Ellis *et al.* (20) mentioned two kinds of models for formation of a strict boundary of PAR. One is based on genome rearrangement such as insertion, deletion, inversion, and translocation; the model proposed above postulating illegitimate recombination belongs to this category. The other model is based on the following 'attrition' process, and our model is different from the 'attrition' model. High-frequency recombination acts to maintain sequence similarity in the PAR between the sex chromosomes, but this event is infrequent in sequences close to the PAB. The attrition model predicts that when enough sex-specific differences accumulate immediately distal to the boundary, the probability of recombination in the region with mismatches is reduced. Once recombination is limited, divergence between the X and Y chromosomes accumulates more rapidly until recombination events no longer include the region of mismatches, resulting in formation of a new strict boundary. In the attrition model, the region of reduced homology in PABXY1 (i.e., downstream of the *Alu* insertion site) is taken to correspond to the portion derived from a single PABL after attrition occurred. Therefore, when based on this model, the most probable explanation for why the 3' homology terminus between PABX1 and PABY1 is almost identical to the termini among PABLs is a strong functional constraint that preserves the 3'-terminal position during course of evolution. Furthermore, an extra mechanism may act to put an evolutionally stable interface between the regions of high- and reduced-homology (i.e., the *Alu*-insertion site) during attrition. Our explanation for the reduced-homology supposed in Figure 7 is different from this attrition process, and is based mainly on the finding that the diversity between PABX1 and PABY1 is similar to the diversity of either PABX1 or PABY1 from PABLs in the respective region (Fig. 5A). In this connection, it is worthwhile to consider the sequence organization of the

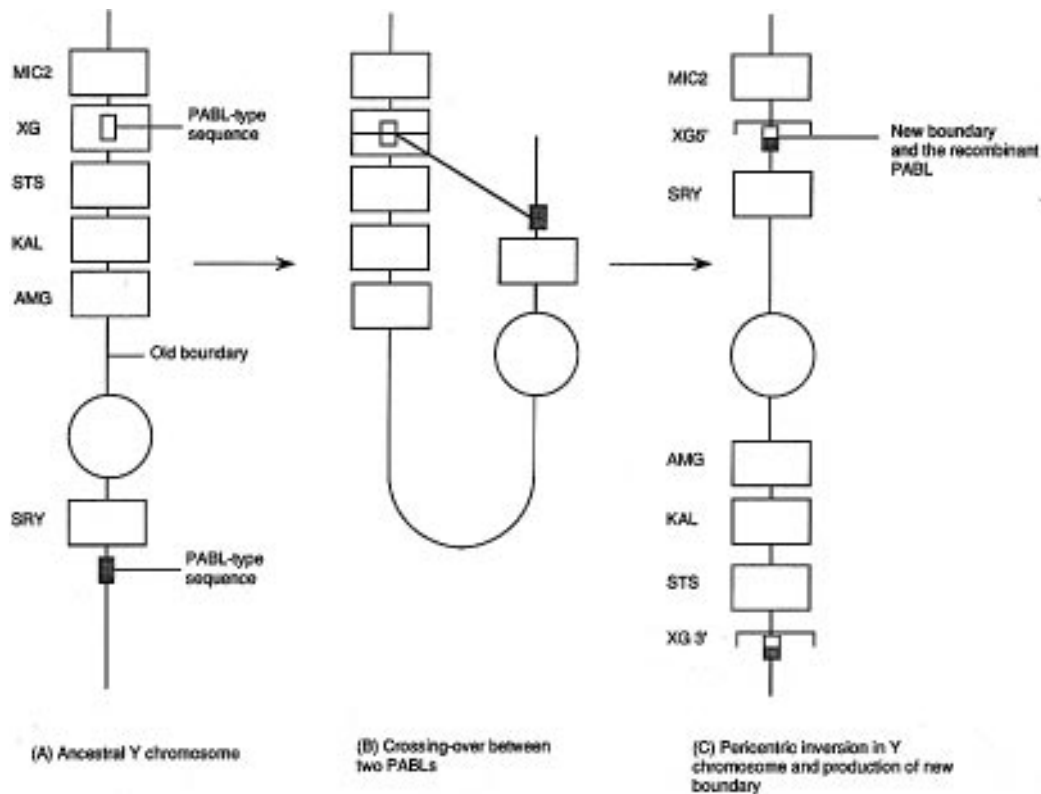


Figure 7. Model of the evolutionary formation of the PABY1, in a modification of the model proposed by Ellis *et al.* (31). As the molecular process in the pericentric inversion hypothesized by them, an illegitimate recombination between two PABLs is postulated. (A) According to Ellis *et al.*, in early primates, PAB1 was somewhere proximal to the present-day amorgenin gene on the X chromosome (AMGX). We suggest that, in the old Y chromosome, one PABL was within XG and another was 5 kb distal to SRY. (B) After the divergence of lineages to higher primates and prosimians, a new boundary is considered to have been formed by a pericentric inversion on the Y chromosome initiated by break points inside XG and 5 kb distal to SRY. As the molecular process of this inversion, we propose illegitimate recombination between the two PABLs mentioned above. (C) An inverted Y chromosome with the present PABY1 was thus formed. Further evolutionary events such as pericentric inversion transferring AMGY to its present location were described by Ellis *et al.* (31).

boundary of PAR2, the pseudoautosomal regions of the long arms of the human sex chromosomes. At the breakpoint of the X-Y sequence homology in PAR2, Kvaløy *et al.* (18) found a portion of LINE-1 sequence (about 780 nt), and they proposed that illegitimate recombination between two independent LINES on earlier X and Y chromosomes were involved in the formation of the present PAR2 boundary, PABXY2. The similarity between the reported LINE-1 sequences in PABX2 and PABY2 is 92% nucleotide identity. One explanation for this reduced homology may be due to attrition derived from a single LINE-1 sequence. To test this possibility, we used 780 nt of the LINE-1 sequences of X and Y chromosomes in a search of GenBank sequences. LINE-1 of the X chromosome had 98% identity with LINE-1 of HSL1G and LINE-1 of the Y chromosome had 96% identity with LINE-1 of HSRETBLAS, indicating separate origins for the LINE-1 sequences in PABX2 and PABY2. This finding is not consistent with the attrition model postulating these two LINE-1 sequences with 92% identity have been derived from a single LINE-1, but consistent with the model supposing the illegitimate recombination between two LINES. Boundaries of both PAR1 and PAR2 therefore were probably produced by analogous processes, that is, by illegitimate recombination between repetitive elements: PABLs for PABXY1 and LINES for PABXY2. Analogous processes may have been involved in forming the present-day human genome, which is composed of GC% mosaic structures.

MATERIALS AND METHODS

Cloning and sequencing

A cosmid library constructed from the total human DNA of an HLA homogeneous B cell line on vector pWE15 by Inoko *et al.* (34) and a λ -EMBL3 library constructed from the human DNA of peripheral blood cells (35) were used. Six human cDNA libraries cloned on vector λ gt10 or λ gt11 were obtained from Clontech (Palo Alto, CA): placenta 5'-stretch plus (λ gt11), placenta (λ gt11), monocyte (λ gt11), B cell (λ gt10), spleen (λ gt10), and skin fibroblast (λ gt10). Cloning of PABLs, their subcloning into pUC118, sequencing, and database searches of GenBank/EMBL/DDBJ, as well as of PIR and Swiss-Prot protein databases, were done as described previously (13,28).

Sequence alignment and construction of phylogenetic trees

Alignments and calculations of sequence identity were done with the MALIGN program available on DDBJ. For identification of the core sequence of PABLs, all pairs of sequences with PABLs or PABXY1 were first aligned, and multiple alignments of the similar portions thus found were done as described by Hein (36). Phylogenetic trees were constructed by using the neighbor-joining method (37). Evolutionary rates were estimated as described by Nei (38).

ACKNOWLEDGEMENTS

The authors are very grateful to Dr N. Ellis for kindly providing us with his unpublished sequence and valuable comments; to Drs K. Sugaya, T. Tenzen and K. Matsumoto for discussion; to Mrs Y. Miyauchi for technical assistance. The computers at the DDBJ and the Human Genome Center of Japan were used. This work was supported by a JSPS Fellowship for Japanese Junior Scientists to T.F., Grant-in-Aid for Creative Basic Research (Human Genome Project) to T.I. and for Scientific Research on Priority Areas (Genome Informatics) to T.I., and by Grants-in-Aid of Scientific Research from the Ministry of Education, Science and Culture of Japan to T.F. and T.I.

ABBREVIATIONS

MHC, major histocompatibility complex; PAB, pseudoautosomal boundary; PABL, pseudoautosomal boundary-like sequence; PABX1, pseudoautosomal boundary sequence of short arm of X chromosome; PABY1, pseudoautosomal boundary sequence of short arm of Y chromosome; PABXY1, PABX1 and PABY1 sequences; PAR1, pseudoautosomal region of short arm of sex chromosomes; PAR2, pseudoautosomal region of long arm of sex chromosomes; EST, expressed sequence tag.

REFERENCES

- Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F. (1985) The mosaic genome of warm-blooded vertebrates. *Science*, **228**, 953–958.
- Ikemura, T. (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol. Biol. Evol.*, **2**, 13–34.
- Aota, S. and Ikemura, T. (1986) Diversity in G+C content at the third position of codons in vertebrate genes and its cause. *Nucleic Acids Res.*, **14**, 6345–6355.
- Bernardi, G. (1989) The isochores organization of the human genome. *Annu. Rev. Genet.*, **23**, 637–661.
- Bernardi, G. (1993) The isochores organization of the human genome and its evolutionary history—A review. *Gene*, **135**, 57–66.
- Ikemura, T. and Aota, S. (1988) Global variation in G+C content along vertebrate genome DNA; possible correlation with chromosome band structures. *J. Mol. Biol.*, **203**, 1–13.
- Korenberg, J. R. and Rykowski, M. C. (1988) Human genome organization: *Alu*, Lines and the molecular structure of metaphase chromosome bands. *Cell*, **53**, 391–400.
- Gardiner, K., Aissani, B. and Bernardi, G. (1990) A compositional map of human chromosome 21. *EMBO J.*, **9**, 1853–1858.
- Ikemura, T., Wada, K. and Aota, S. (1990) Giant G+C% mosaic structures of the human genome found by arrangement of GenBank human DNA sequences according to genetic positions. *Genomics*, **8**, 207–216.
- Ikemura, T. and Wada, K. (1991) Evident diversity of codon usage patterns of human genes with respect to chromosome banding patterns and chromosome numbers; relation between nucleotide sequence data and cytogenetic data. *Nucleic Acids Res.*, **19**, 4333–4339.
- Holmquist, G. P. (1992) Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.*, **51**, 17–37.
- Craig, J. M. and Bickmore, W. A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nature Genet.*, **7**, 376–382.
- Fukagawa, T., Sugaya, K., Matsumoto, K., Okumura, K., Ando, A., Inoko, H. and Ikemura, T. (1995) A boundary of long-range G+C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence exists near the boundary. *Genomics*, **25**, 184–191.
- Cooke, H. J., Brown, W. R. and Rappold, G. A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature*, **317**, 687–692.
- Simmler, M.-C., Rouyer, F., Vergnaud, G., Nystrom-Lahti, M., Ngo, K. Y., de la Chapelle, A. and Weissenbach, J. (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature*, **317**, 692–697.
- Ellis, N. and Goodfellow, P. N. (1989) The mammalian pseudoautosomal region. *Trends Genet.*, **5**, 406–410.
- Rappold, G. A. (1993) The pseudoautosomal regions of the human sex chromosomes. *Hum. Genet.*, **92**, 315–324.
- Kvaloy, K., Galvagni, F. and Brown, W. R. A. (1994) The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Hum. Mol. Genet.*, **3**, 771–778.
- Ellis, N. A., Goodfellow, P. J., Pym, B., Smith, M., Palmer, M., Frischauf, A.-M. and Goodfellow, P. N. (1989) The pseudoautosomal boundary in man is defined by an *Alu* repeat sequence inserted on the Y chromosome. *Nature*, **337**, 81–84.
- Ellis, N., Yen, P., Neiswanger, K., Shapiro, L. J. and Goodfellow, P. N. (1990) Evolution of the pseudoautosomal boundary in Old World monkeys and great apes. *Cell*, **63**, 977–986.
- Uberbacher, E. C. and Mural, R. J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA*, **88**, 11262–11265.
- Li, W.-H., Luo, C.-C. and Wu, C.-I. (1985) In MacIntyre, R. J. (ed.), *Molecular Evolutionary Genetics*. Plenum Press, New York, pp. 1–94.
- Senger, G., Ragoussis, J., Trowsdale, J. and Sheer, D. (1993) Fine mapping of the human MHC class II region within chromosome band 6p21 and evaluation of probe ordering using interphase fluorescence *in situ* hybridization. *Cytogenet. Cell Genet.*, **64**, 49–53.
- Mandel, J.-L., Monaco, A. P., Nelson, D. L., Schlessinger, D. and Willard, H. (1992) Genome Map III. *Science*, **258**, 87–102.
- Whitfield, L. S., Hawkins, T. L., Goodfellow, P. N. and Sulston, J. (1995) 41 kilobases of analyzed sequence from the pseudoautosomal and sex-determining region of short arm of human Y chromosome. *Genomics*, **27**, 306–311.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge Univ. Press, Cambridge, UK.
- Selig, S., Okumura, K., Ward, D. C. and Cedar, H. (1992) Delineation of DNA replication time zones by fluorescence *in situ* hybridization. *EMBO J.*, **11**, 1217–1225.
- Sugaya, K., Fukagawa, T., Matsumoto, K., Mita, K., Takahashi, E., Ando, A., Inoko, H. and Ikemura, T. (1994) Three genes in the human MHC class m region near the junction with the class III: gene for receptor of advanced glycosylation end products, PBX2 homeobox gene and a Notch-homolog, human counterpart of mouse mammary tumor gene int-3. *Genomics*, **23**, 408–419.
- Goodfellow, P., Pym, B., Mohandas, T. and Shapiro, L. J. (1984) The cell surface antigen locus, MIC2X, escapes X-inactivation. *Am. J. Hum. Genet.*, **36**, 777–782.
- Schiebel, K., Weiss, B., Wohrle, D. and Rappold, G. (1993) A human pseudoautosomal gene, ADP/ATP translocase, escapes X-inactivation whereas a homologue on Xq is subject to X inactivation. *Nature Genet.*, **3**, 82–87.
- Ellis, N. A., Ye, T.-Z., Patton, S., German, J., Goodfellow, P. N. and Weller, P. (1994) Cloning of PBDX, an MIC2-related gene that spans the pseudoautosomal boundary on chromosome Xp. *Nature Genet.*, **6**, 394–400.
- Ellis, N. A., Tippet, P., Petty, A., Reid, M., Weller, P. A., Ye, T. Z., German, J., Goodfellow, P. N., Thomas, S. and Banting, G. (1994) PBDX is the XG blood group gene. *Nature Genet.*, **8**, 285–290.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J. and Willard, H. (1992) The human XIST gene: Analysis of 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell*, **71**, 527–542.
- Inoko, H., Ando, A., Kimura, M. and Tsuji, K. (1985) Isolation and characterization of the cDNA clone and genomic clones of a new HLA class II antigen heavy chain, DO alpha. *J. Immunol.*, **135**, 2156–2159.
- Tomatsu, S., Kobayashi, Y., Fukumaki, T., Yubisui, T., Orii, T. and Sakaki, Y. (1989) The organization and the complete nucleotide sequence of human NADH-cytochrome b5 reductase gene. *Gene*, **80**, 353–361.
- Hein, J. (1990) Unified approach to alignment and phylogenies. *Methods Enzymol.*, **183**, 626–645.
- Saitou, N. and Nei, M. (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
- Nei, M. (1987) *Molecular Evolutionary Genetics*, Columbia Univ. Press, New York.
- Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-PhOH-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.
- Sneath, P. H. A. and Sokal, R. R. (1973) *Numerical taxonomy*. Freeman, San Francisco.
- Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, **16**, 111–120.