 Human Molecular Genetics, 1997, Vol. 6, No. 7 **991–1002**

ARTICLE

Interchromosomal duplications of the adrenoleukodystrophy locus: a phenomenon of pericentromeric plasticity

Evan E. Eichler[1,*], Marcia L. Budarf[2], Mariano Rocchi[3], Larry L. Deaven[4], Norman A. Doggett[4], Antonio Baldini[5], David L. Nelson[5] and Harvey W. Mohrenweiser[1]

[1]Human Genome Center, Biology and Biotechnology Research Program, L-452, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA, [2]Children's Hospital of Philadelphia, 34th Street and Civic Center Blvd., Philadelphia, PA 19104, USA, [3]Istituto di Genetica, Via Amendola 165/A70126, Bari, Italy, [4]Life Sciences Division and Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA and [5]Department of Molecular and Human Genetics, Human Genome Center, Baylor College of Medicine, Houston, TX 77030, USA

Received February 25, 1997; Revised and Accepted April 21, 1997

A 9.7 kb segment encompassing exons 7–10 of the adrenoleukodystrophy (*ALD*) locus of the X chromosome has duplicated to specific locations near the pericentromeric regions of human chromosomes 2p11, 10p11, 16p11 and 22q11. Comparative sequence analysis reveals 92–96% nucleotide identity, indicating that the autosomal *ALD* paralogs arose relatively recently during the course of higher primate evolution (5–10 million years ago). Analysis of sequences flanking the duplication region identifies the presence of an unusual GCTTTTTGC repeat which may be a sequence-specific integration site for the process of pericentromeric-directed transposition. The breakpoint sequence and phylogenetic analysis predict a two-step transposition model, in which a duplication from Xq28 to pericentromeric 2p11 occurred once, followed by a rapid distribution of a larger duplicon cassette among the pericentromeric regions. In addition to facilitating more effective mutation detection among ALD patients, these findings provide further insight into the molecular basis underlying a pericentromeric-directed mechanism for non-homologous interchromosomal exchange.

INTRODUCTION

Adrenoleukodystrophy (ALD) is a relatively common X-linked neurodegenerative disorder with an estimated incidence among males of 1/15 000–1/20 000 (MIM #300100, McKusick). The disease, which exists primarily in two forms, cerebral childhood ALD and adrenomyeloneuropathy among adults, is characterized by accumulation of saturated very long chain fatty acids in serum and other tissues, adrenal insufficiency and a general demyelination of the central nervous system. The molecular basis of this disease was determined by the identification of the *ALD* gene in Xq28 (1). This 21 kb gene consists of 10 exons encoding a peroxisomal-specific ATP-binding cassette transporter (ABC) protein (1,2) whose dysfunction leads to aberrant methylation and/or aberrant transport of very long chain fatty acids. Several systematic analyses of ALD kindreds have revealed that ~90% of molecular lesions associated with the disease are the result of nonsense, missense or frameshift point mutations within the coding portion of the *ALD* gene (3–11). Unambiguous detection of mutation in *ALD*, however, has been confounded by the presence of other cross-hybridizing genomic fragments corresponding to the distal portion of the *ALD* gene, but not localized to the X chromosome (2,3,5,10). Based on the pattern of cross-hybridization, it was purported that these paralogous *ALD* copies represented non-functional pseudogenes (2,5). One such pseudogene has been identified recently (10), which raises the possibility that co-amplification of pseudogene products could present serious impediment to molecular diagnosis of ALD.

The regions flanking the Xq28 *ALD* locus demonstrate an unusual degree of genomic instability (12,13). A cluster of tandem and polymorphic duplications of the red and green cone pigment genes (*RCP* and *GCP*) is situated ~600 kb telomeric to the *ALD* locus (2,13). These genes exhibit a high degree of heteromorphism, with variable numbers of copies of these genes correlating with differential red–green color vision sensitivity among males. A region of Xq28, 20 kb proximal of the *ALD* locus, has been identified recently which shows an unusual proclivity to duplicate and transpose to the pericentromeric regions of chromosomes (12). Analysis of the region revealed that a 26.5 kb segment had been transposed recently to human cytogenetic band 16p11.1. The duplicon included the entire creatine transporter (*SLC6A8*) and part of the *CDM* (*DXS1357E*) genes (GenBank accession nos U41302 and U36341). In

*To whom correspondence should be addressed. Tel: +1 510 423 7831; Fax: +1 510 422 2282; Email: eichler1@llnl.gov

addition, cytogenetic mapping among non-human higher primates indicated that the region had been duplicated multiple times to the pericentromeric regions of additional autosomes, without synteny to human chromosome 16 (12). These data suggest that an ~700 kb interval extending from *SLC6A8* to the *RCP*/*GCP* region of the X chromosome had been subjected to signficant interchromosomal and intrachromosomal duplications in the last 5 million years.

During our analysis of the creatine transporter (*CTR*)–*CDM* duplicon, we identified a second domain of paralogy which included a substantial portion of the *ALD* gene. Preliminary analysis indicated that this region had been duplicated independently and also with a strong transposition bias toward the pericentromeric regions of specific chromosomes, particularly cytogenetic band 2p11.1 (12). These findings were consistent with the existence of additional fragments, not linked to the X chromosome, which cross-hybridized with *ALD* cDNA probes (2). We sought to investigate the nature of these duplications in more detail for two reasons. Determination of the size, timing and breakpoint sequence of the *ALD* duplications would provide greater insight into the phenomenon of pericentromeric-directed transposition and the propensity of this region of Xq28 to duplicate and transpose in human evolution. An understanding of the architecture and sequence of *ALD* paralogs would also facilitate mutation detection by enabling more effective design of primers to analyse *bona fide* Xq28 sequence.

RESULTS

Identification and mapping of *ALD* duplications

Using cosmids spanning a 50 kb interval between the *CTR* and *ALD* genes as probes, we previously reported strong fluorescent *in situ* hybridization (FISH) cross-hybridizing signals at human cytogenetic bands 2p11, 16p11 and Xq28 (12). Once the *CTR*/*CDM* paralogy domain had been defined precisely at the sequence level, FISH analysis was repeated to analyze the *ALD* paralogy domain in more detail. An X-chromosome-specific cosmid probe, U184E11, was chosen which strongly hybridized to *ALD* cDNAs, H8 and T19, (2) and which largely excluded the *CTR*/*CDM* duplication domain. FISH of human metaphase spreads with U184E11 suggested the presence of at least five *ALD* paralogous segments localized in Xq28, 2p11, 10p11, 16p11 and 22q11 (Fig. 1). PCR screening of a human somatic cell hybrid monochromosomal panel (NIGMS) using oligonucleotide primers specific for *ALD*Xq28 exons 7 and exons 9 (Table 1, products 4 and 11) confirmed these autosomal locations and that the duplications involved the *ALD* locus (data not shown). As a final verification of the existence of the autosomal copies of *ALD*, products 4 and 11 (Table 1) were used as probes to screen flow-sorted chromosome-specific cosmid libraries, LL02NC01, LA10NC02, LA16NC02 and LL22NC03, corresponding to chromosomes 2, 10, 16 and 22, respectively. Cosmid clones 2-11c12, 2-74a10 (from LL02NC01), 10-204a1 (from LA10NC02), 16-341b10, 16-366a10, 16-431f4 (from LA16NC02) and 22-11c7 (from LL22NC03) were identified. Subsequent PCR analysis (Fig. 2) revealed that these cosmid clones represented autosomal duplications of the *ALD* locus.

In order to map more precisely the genomic location of each *ALD* paralog, hamster somatic cell hybrid chromosome-specific deletion and radiation hybrids were screened by PCR using the Xq28-derived primers. The chromosome 10 *ALD* locus was regionally localized by screening a chromosome 10 hybrid panel (14) using products 4 and 11 (Table 1). Hybrids 175, 132, 57, 10 and 43 from this panel were positive for both products by PCR, and hybrids 168 and 170 were negative, which is consistent with a map position for *ALD*10p11 in distal 10p11.1 to proximal 10p11.2. Similarly, the chromosome 22 *ALD* locus was finely localized by screening of chromosome 22 hybrids (15). Hybrids Cl 6-2/EG, GM10888, D655 and Cl-4/GB were positive for products 4 and 11, and Cl1-9/5878 was negative, which suggests a location for *ALD*22q11 proximal to the commonly deleted region (cytogenetic band 22q11.2). FISH analysis using *ALD*-specific cosmids as probes against human metaphase spreads of velocardiofacial/DiGeorge (VCF/DG) patients confirmed this location (data not shown). The map position of the chromosome 16 *ALD* locus was resolved by screening a CEPH mega-YAC panel spanning the 16p11.1–11.2 interval (16). Probes 4 and 11 (Table 1) were both positive for My662d12. My662d12 is located ~500 kb distal to YAC My895g9, which contains the *CTR–CDM* duplicon (12). The map position of the *ALD* paralog on chromosome 2 was not refined due to the unavailability of well-characterized hybrids or YAC contig maps in this region.

Table 1. *ALD* PCR probes. PCR probes are numbered sequentially relative to the orientation of *ALD* transcription of Xq28

PCR Product	Location	Length (bp)	PCR Annealing	Primer	Primer Sequence
1	intron 6	186	60	86474	51868-51891
				86475	*52054-52031
2	exon 6	298	60	73273	53050-53063
				73274	*53316-53348
3	intron6	246	60	86476	53296-53320
				86477	*53518-53542
4	exon 7	219	60	73275	53581-53605
				73276	*53776-53800
5	exon-intron 7a	809	60	73275	53581-53605
				83184	*54365-54390
6	intron 7a	489	65	83183	53901-53926
				83184	*54365-54390
7	intron 7b	1479	65	83183	53901-53926
				83186	*55355-55380
8	intron 7	369	65	83185	55011-55035
				83186	*55355-55380
9	intron-exon 8	1159	60	83185	55011-55035
				79492	*56145-56170
10	exon 8	179	60	79491	55991-56014
				79429	*56145-56170
11	exon 9	198	60	73277	56220-56254
				73278	*56393-56418
12	exon 9-exon 10	606	60	73277	56220-56254
				79494	*56801-56826
13	exon 9-IG2	4737	60	73277	56220-56254
				83190	*59913-59946
14	exon 10	440	60	79493	56496-56521
				79494	*56801-56826
15	IG1	356	60	83187	57601-57625
				83188	*57821-57846
16	IG1-IG2	2345	60	83187	57601-57625
				83190	*59913-59946
17	IG2	245	60	83189	59701-59726
				83190	*59913-59946
18	IG3	224	60	83191	62571-62594
				83192	*62771-62795
19	IG4	402	60	110516	63169-63192
				110517	*63548-63571
20	IG5	359	55	110518	63871-63892
				110519	*65206-64230

The length, location and annealing temperature for each primer are indicated. The sequence of each primer is shown with respect to GenBank accession no. U52111. IG denotes intergenic PCR products. *Indicates that the reverse complement of this sequence was used to generate an oligonucleotide primer.

Extent of paralogy

In order to assess the size of the *ALD* duplicons on various chromosomes, 23 sets of Xq28-derived primer pairs were

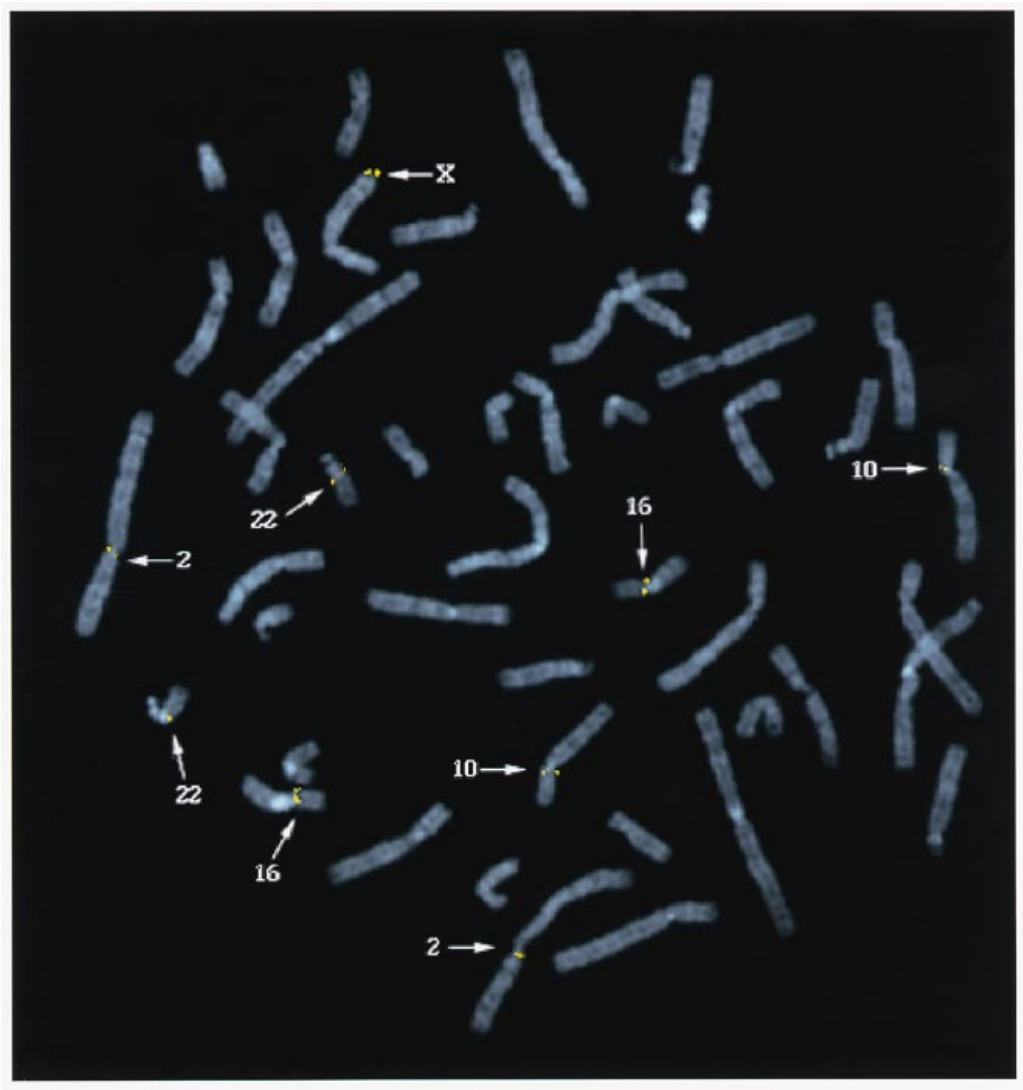

Figure 1. Cytogenetic mapping of *ALD* cosmid U184E11. Chromosome metaphase spreads were prepared from peripheral lymphocytes from a male donor and hybridized with biotin-labeled, nick-translated cosmid probe, U184E11. Cosmid U184E11 was derived from an X chromosome-specific library and includes the entire *ALD* gene (12). Hybridization signals to Xq28, 2p11, 10p11, 16p11 and 22q11 are indicated by arrows.

developed, extending from *CDM* exon 5 (position 16 851, GenBank U52111) to the first exon of a putative plexin-related gene (position 79 726). A subset of these and their positions with respect to the *ALD* intron–exon structure is shown in Table 1 and Figure 2. Each cosmid was subjected to PCR analysis using a battery of these primer pairs (Fig. 2) in order to define the extent of paralogy. Occasionally, cosmid vector–insert junctions occurred within duplicated portion of the *ALD* locus. To eliminate the possibility of this artifact, all Xq28–autosome boundaries were confirmed by PCR using monochromosomal panel DNAs as templates (data not shown). In addition, all breakpoint junctions were sequenced and compared with Xq28 *ALD* genomic sequence (GenBank U52111), confirming the end of Xq28 paralogy (see below). The data indicated that the autosomal duplications were all similar in size (~10 kb) and encompassed exons 7–10 of the *ALD* genomic structure (Fig. 2). The genomic organization of the *ALD* cassettes was highly conserved among the four autosomal loci (Fig. 2). The absence of the proximal portion of the *ALD* gene indicates that the autosomal copies represent truncated non-processed pseudogenes.

Comparative sequence and phylogenetic analyses

Genomic fragments corresponding to exons 8, 9 and 10 were sequenced for the four autosomal paralogs of the *ALD* locus and deposited in GenBank (accession nos U90288, U90289, U90290 and U90291 for *ALD*10p11, *ALD*16p11, *ALD*2p11 and *ALD*22q11, respectively). The genomic sequences corresponding to exons 8, 9 and 10 were aligned with previously published Xq28 *ALD* sequence (Fig. 3). An average nucleotide identity of 94.6% was calculated over this 790 bp region for the five *ALD* paralogs (Table 2). *ALD*2p11 exhibits the greatest divergence, with an average nucleotide identity of 93.3%. *ALD*16p11 and *ALD*22q11 exhibit the least divergence, with an average nucleotide identity of 95.3% to all other *ALD* paralogs.

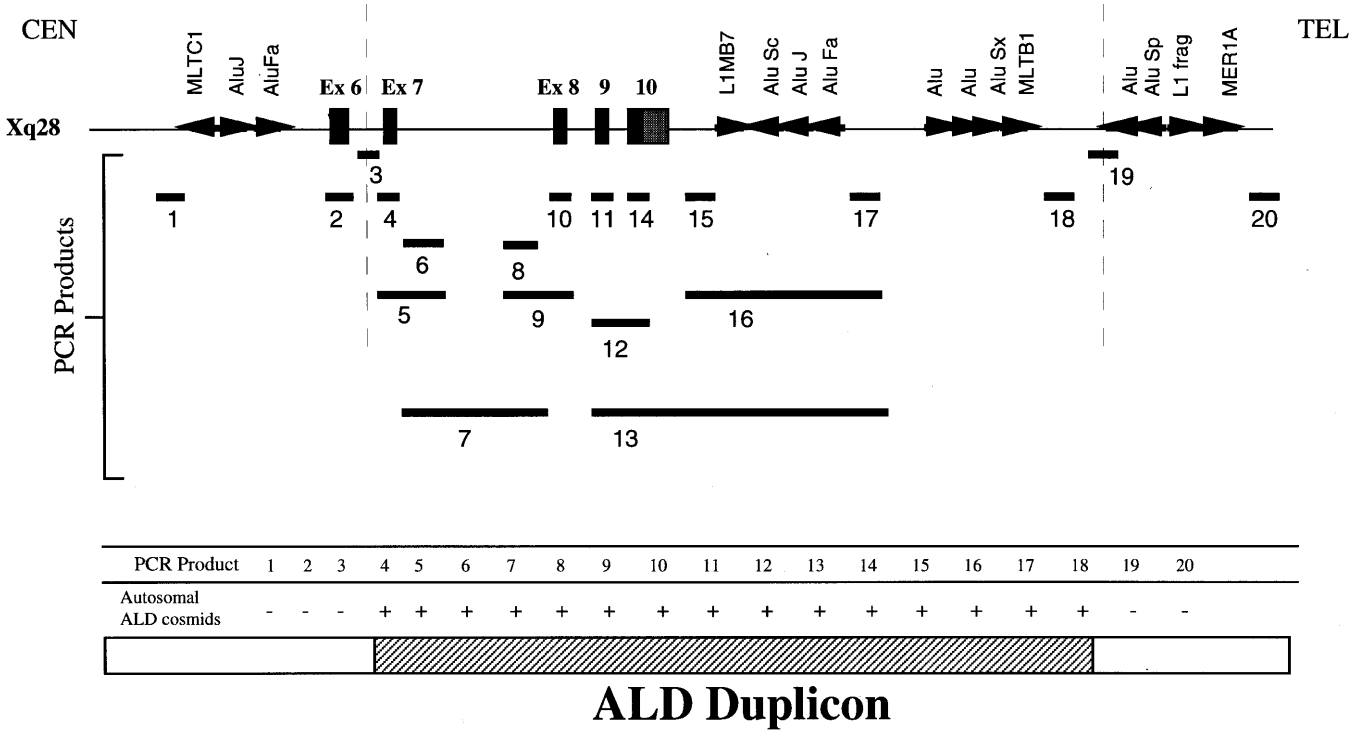

Figure 2. The extent of *ALD* paralogy. The exon–intron structure of the distal portion of the Xq28 ALD gene is shown, as derived from GenBank U52111. Filled boxes indicate the position of exons (6–10). The shaded box shows the position of the 3′ UTR of the *ALD* gene. Various repetitive structures (SINEs and LINEs) and their orientation are shown by arrows. Horizontal bars show the position of PCR products used in determining the organization of the various *ALD* duplicons (see Table 1). Bars which overlap share one common oligonucleotide primer. PCR analysis, using primer pairs from Table 1, was performed using autosome-specific *ALD* cosmids as templates (see Materials and Methods). + indicates that a product of approximately the same size as expected (based on Xq28 sequence) was generated.

Table 2. Pairwise nucleotide identity among *ALD* paralogs

ALD paralog	2p11	10p11	16p11	22q11	Xq28
2p11	–	93.2	92.4	94.5	93.2
10p11	–	–	94.5	95.4	95.9
16p11	–	–	–	95.6	95.8
22q11	–	–	–	–	96.1
Xq28	–	–	–	–	–

A 790 bp genomic segment corresponding to *ALD* exons 7–10 was aligned as indicated in Figure 3. The percentage nucleotide identity for each pairwise alignment was calculated using the BestFit program (GCG software).

Maximum parsimony analysis was employed in an attempt to reconstruct the phylogeny of the *ALD* duplications. Parsimony analysis without a defined ancestral state generates two equally parsimonious trees, one of which is slightly favored by bootstrap analysis (Fig. 4) (100 branch and bound replicates, tree length = 105, CI = 0.94). This analysis supports the existence of a clade occupied by sequences from *ALD*22q11, *ALD*2p11 and *ALD*16p11, which is separated from *ALD*10p11 and the functional Xq28 duplicon by five unique genetic mutational events (Fig. 4). Within this clade, a remarkable asymmetry is observed, with 34 genetic events separating *ALD*2p11 from its nearest node. A majority-rule consensus tree of both equally parsimonious trees was uninformative (data not shown). The majority-rule consensus tree and bootstrap analysis support a model of rapid duplication and dissemination of *ALD* copies over a short period of recent human evolution, as shown by the ambiguity of branchpoint resolution (bootstrap values of ~50% or less).

ALD Xq28–autosome paralogy breakpoints

PCR-generated products derived from the Xq28 sequence (Table 1) were used as probes to refine the location of the junction fragments between Xq28 and autosomal *ALD* copies. Southern blot analysis of digested *ALD* paralogous cosmids using these products as probes (data not shown) resolved both breakpoints to within 50 bp. Sequence analysis of both Xq28–autosome breakpoints indicated that the junctions were identical for all four autosomal duplicons. The original 5′ transposition breakpoint (relative to the orientation of Xq28 *ALD* transcription) lies within intron 6 of the *ALD* gene. Located <100 bp proximal to the *ALD* junction sequence, an unusual GCTTTTTGC repeat structure was observed among the autosomal *ALD* copies (Figs 5a and 6). The sequence consists of GCTTTTTGC direct, non-tandem repeats. These elements occur with a density of one repeat motif every 30 bp. BLAST searches of this sequence against GenBank databases (release 1.4.9) revealed a strong similarity of this repeat sequence (~70% over 100 bp) to a previously described motif implicated in the transposition of the chromosome 2p11.1 immunoglobulin light chain Vκ genes to chromosome 1 (17). The Xq28 3′ junction sequence occurs ~4 kb distal to the last exon of the *ALD* gene, near a cluster of inverted Alu repeats (Figs 2 and 7). An Alu monomer repeat sequence was identified precisely at the 3′ breakpoint junction among all four autosomal *ALD* duplicons (Fig. 5b). Sequence similarity among the

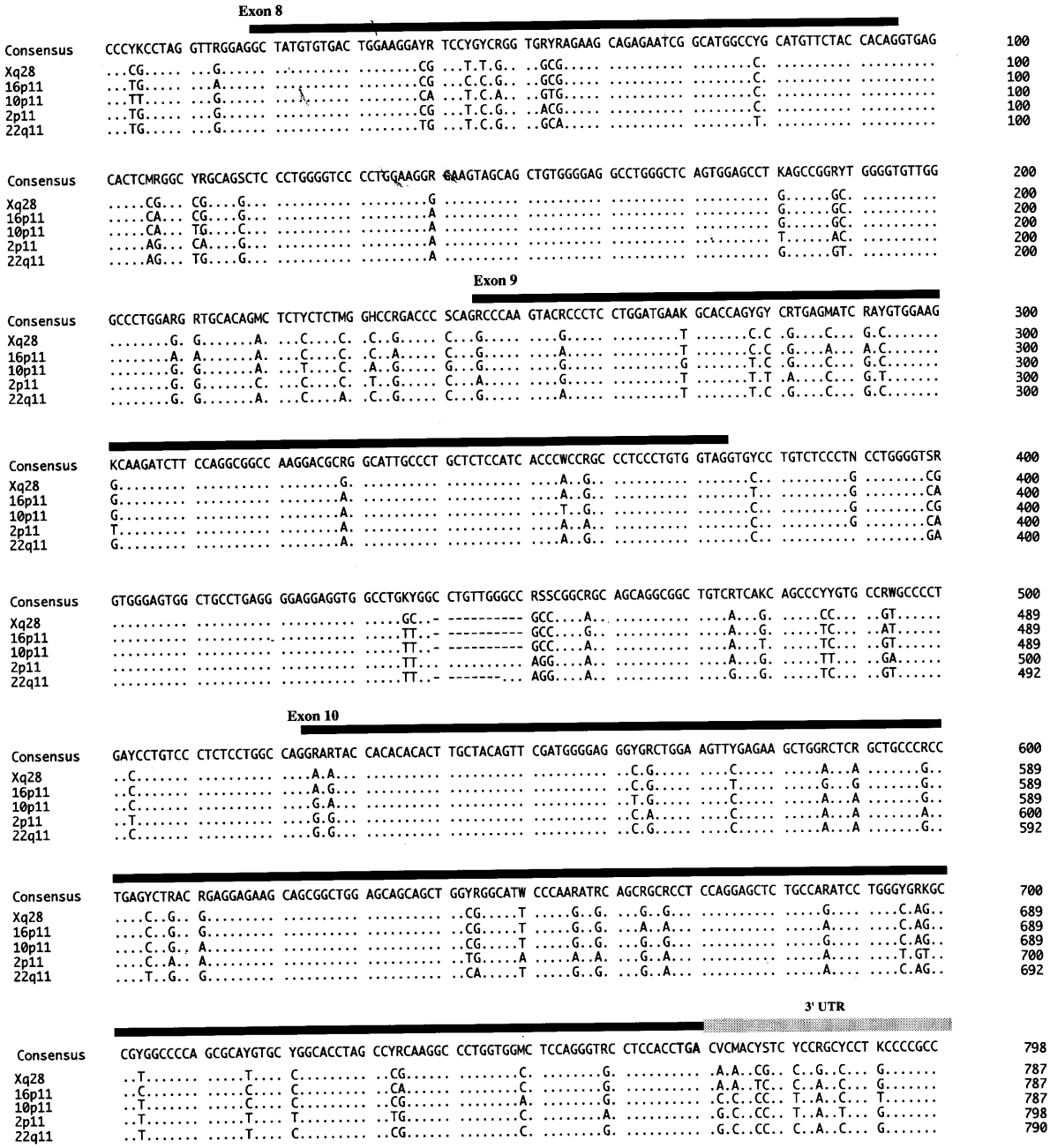

Figure 3. *ALD* paralogous sequence alignment. A ~790 bp genomic fragment corresponding to exons 8–10 was sequenced from the various chromosome-specific *ALD* cosmids (accession nos U90288, U90289, U90290 and U90291 for cosmids 10-204a1, 16-366a10, 2-11c12 and 22-11c7, respectively). The corresponding Xq28 genomic sequence was derived from GenBank accession no.U52111. Sequences were aligned using the BestFit program (GCG software) and a consensus sequence was extracted. Divergent nucleotides are indicated among the various *ALD* paralogs. Horizontal bars indicate the position of *ALD* exons.

autosomal *ALD* duplicons extends both 5′ and 3′ of the Xq28–autosome paralogy junctions (Fig. 5a and b).

DISCUSSION

Origin of the *ALD* duplications

We have documented the existence of five copies of the *ALD* locus situated at cytogenetic bands 2p11, 10p11, 16p11, 22q11 and Xq28. The only known functional transcript of *ALD* is derived from the Xq28 locus (1). The *ALD* duplicons at 2p11, 10p11, 16p11 and 22q11 are truncated non-processed pseudogenes. This fact and the observation that mouse–human comparisons map the *ALD* locus only to the syntenic portion of Xq28 (18) may be taken as strong evidence that the Xq28 locus represents the ancestral copy. Comparative sequence analysis between the Xq28 and autosomal loci (Fig. 3 and Table 2) reveals an average nucleotide identity of 94.6%, suggesting that the duplications occurred relatively recently in human evolution. Based on estimates of rates of divergence among paralogous

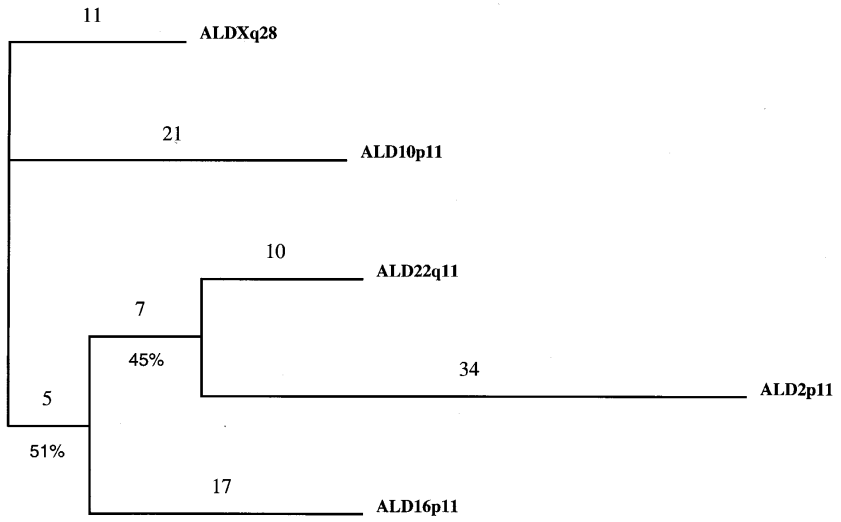

Figure 4. Phylogenetic analysis of *ALD* paralogous sequence. Parsimony analysis of aligned sequences (Fig. 3) generated two equally parsimonious trees. The cladogram which was favored by bootstrap analysis (n = 100 replicates, tree length = 107) is depicted. The number of informative character states separating each sequence is shown above each branch line. The percentages indicate the frequency of bootstrap replicates which support each node.

sequences, which may be as high as 13×10^{-9} mutations per site per year (19), we calculate that the duplications from Xq28 occurred 5–3 million years ago. Parsimony analysis reveals that the majority of the duplications occurred over a very narrow window of evolutionary time, as demonstrated by the ambiguity of branchpoint discernment (Fig. 4). Interestingly, *ALD*2p11 shows the greatest level of divergence (Table 2, Fig. 3). This may indicate that the *ALD*2p11 copy arose earliest (>5 million years ago), or that this particular region of the chromosome is subject to higher rates of mutation.

Pericentromeric-directed transposition

FISH analysis indicates that transposition of Xq28 *ALD* sequence has been directed to the pericentromeric regions of 2p11, 10p11, 16p11 and 22q11 (Fig. 1). Attempts to refine the map locations using deletion and radiation hybrids as well as local maps of contiguous YAC clones confirm an extreme transpositional bias. The *ALD* duplicons appear to map near the heterochromatin–euchromatin boundary (p11.1–p11.2) for chromosomes 2, 10 and 16. *ALD*22q11 is exceptional in this regard as the chromosome is acrocentric and the duplicon maps to 22q11.1–q11.2. Cosmid end sequencing of a chromosome 16-derived *ALD* cosmid (16–341b10, Fig. 2) has identified satellite III sequences (data not shown). Satellite III sequences are commonly associated with the heterochromatic regions near centromeres (20). Mapping and sequencing data would suggest, then, that the *ALD* duplications have been directed to precise locations in the genome near heterochromatic repeat sequences, with a bias to the short arms of specific chromosomes.

Several recent reports indicate that pericentromeric-directed transposition may be a general phenomenon for interchromosomal duplication among the genomes of higher primates (12,21–23). Interchromosomal duplications ranging in size from 10 to 30 kb have been documented for the *CTR* gene (12), immunoglobulin Vκ light chain locus (17,24), the immunoglobulin heavy chain V_H region (25), the *MS29* locus (26) and the neurofibromatosis type 1 (*NF1*) gene (22). All of these recent duplications demonstrate a preference for integration near the pericentromeric regions of chromosomes, suggesting that the organization/structure of these regions of chromosomes may be particularly amenable to site-specific integrations. It has been suggested that this unusual bias may be related to the high degree of homology observed among α-satellite sequences within a given suprachromosomal family (22). This model was based on the observation that six out of seven *NF1* pericentromeric duplications involved chromosomes belonging to the same α-satellite suprachromosomal family (family 2) (22). Our analysis of *ALD* duplications, however, does not support this model. *ALD* transposon integration sites occur on four chromosomes belonging to two distinct α-satellite suprachromosomal families (suprachromosomal families 1 and 2) (27). The pericentromeric bias, thus, cannot be explained by α-satellite sequence homology alone.

GCTTTTTGC: a transposition-associated sequence

The sequence, 100 bp proximal to the 5′ junction of *ALD*2p11, *ALD*10p11, *ALD*16p11 and *ALD*22q11 (as defined by the orientation of transcription of the Xq28 *ALD* gene), revealed the presence of an unusual (A/T)-rich repeat. At the core of the repetitive sequence is a GCTTTTTGC/GCAAAAAGC motif followed by a GC-rich repeat tract ranging in length from 9 to 30 bp (Figs 5a and 7). The sequence and organization of this repeat bears a strong homology to the junction sequence of an immunoglobulin light chain Vκ gene segment transposed from 2p11.1 to chromosome 1 (17,23). Although the ancestral loci of ALD and Vκ on Xq28 and 2p11 are clearly distinct and the events responsible for these transposition events likely occurred independently, the similarities between the ALD and Vκ duplications are striking. Both sets of duplications are estimated to have occurred recently in human evolution (~5 million years ago). Both involved the mobilization of substantial amounts of genomic material (10–15 kb). In addition, both events were directed to the pericentromeric regions of chromosomes (23) and integrated near or at GCTTTTTGC repeats (17). In conjunction

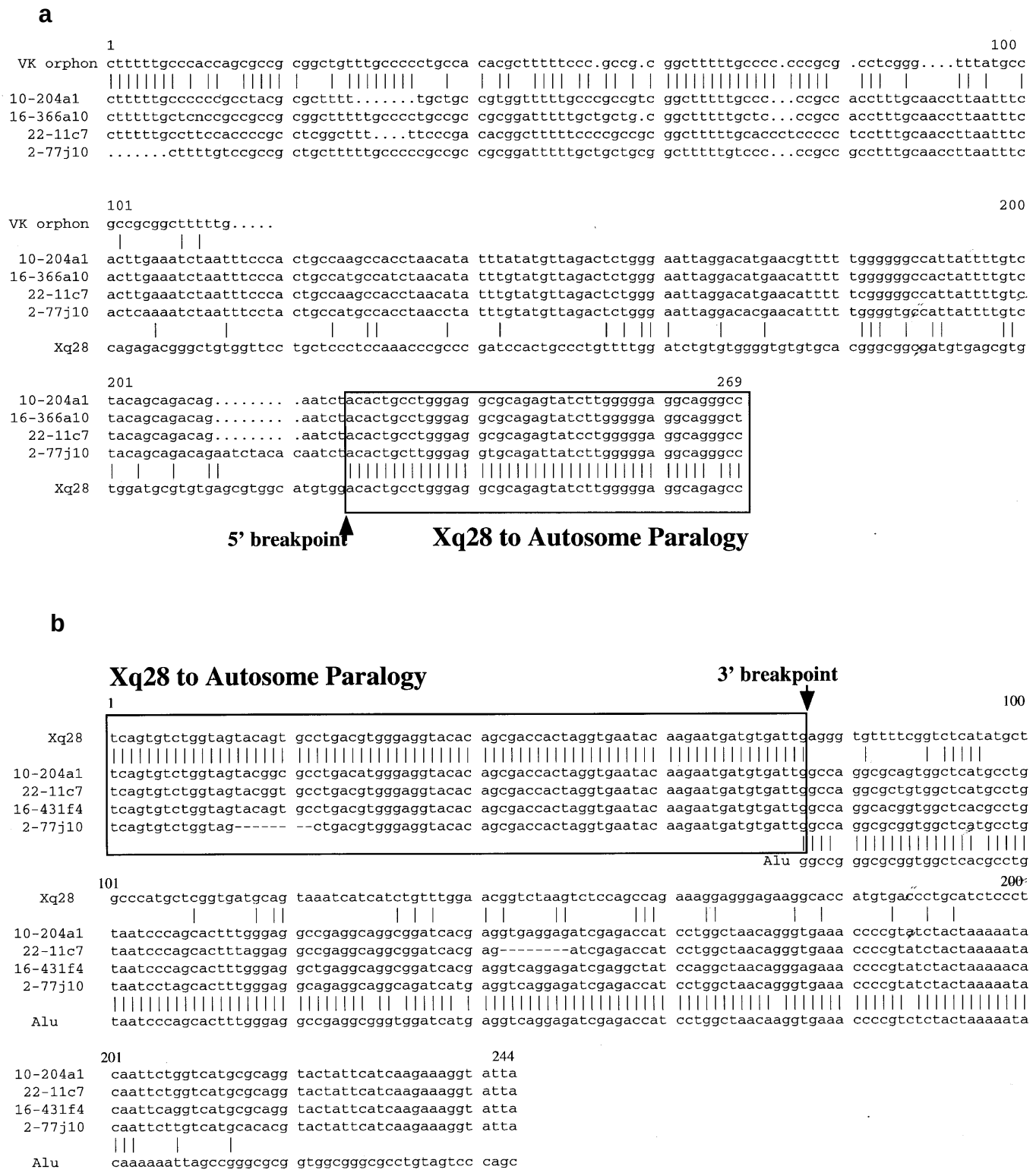

Figure 5. Breakpoint sequence analysis. The sequences flanking the Xq28–autosome *ALD* paralogy breakpoints are shown. The 5′ (**a**) and 3′ (**b**) junctions are determined relative to the orientation of transcription of the *ALD* gene in Xq28 (centromere to telomere). The X–autosome duplicated sequence is boxed, and arrows define the position of the junctions. Vertical lines indicate nucelotide sequence identity between the derived autosomal *ALD* consensus sequence and Xq28/Alu repeat/Vκ orphon sequence. Junction sequences have been deposited in GenBank (5′ breakpoint GenBank accession nos U90295, U90293, U90292, U90294; and 3′ breakpoint accession nos U90296, U90299, U90298 and U90297 for chromosomes 2, 10, 16 and 22, respectively).

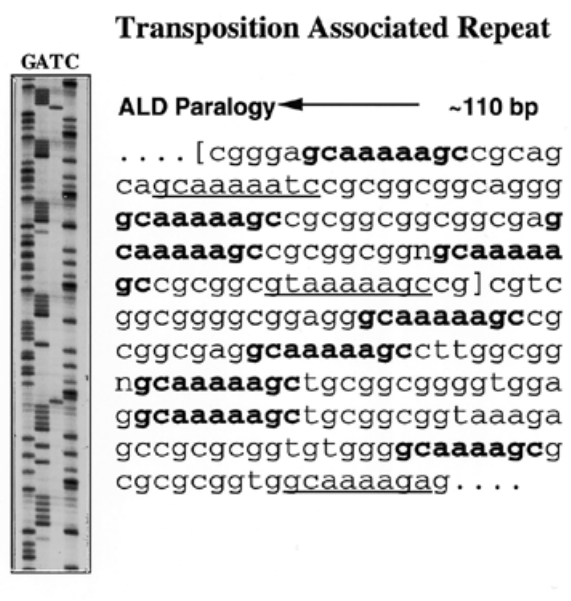

Figure 6. GCTTTTTGC: a transposition-associated sequence. The sequence and organization of the GCAAAAAGC repeat (complementary stand GCTTTTTGC) flanking the 3′ paralogy breakpoint of chromosome 16 cosmid 341b10 are shown. A portion of this sequence from the complementary strand is shown in Figure 5a. Perfect GCAAAAAGC repeat motifs are in bold, while sequences with one base pair degeneracy are underlined. The brackets define the extent of sequence depicted in the gel. The arrow indicates the distance to the 5′ Xq28–autosome junction.

with results from other studies (12,17,25,26), these data would suggest that the human genome has undergone subtle restructuring mediated by a transpositional burst of various genomic segments in the last 5 million years. In addition, GCTTTTTGC repeat sequences appear to have served as sequence-specific integration signals for these transpositions. If such sequences are located preferentially at the pericentromeric boundaries of chromosomes, this may explain the plethora of duplications occurring at these chromosomal positions. In this regard, it should be noted that sequences other than GCTTTTTGC have been identified at the transposition breakpoints in humans (12,17). For example, a CAGGG repeat motif has been identified at the breakpoint of the Xq28 transposed *CTR–CDM* duplicon in 16p11.1 (GenBank accession no. U41302) (12). Although the sequence of this repeat clearly differs from the GCTTTTTGC sequence associated with *ALD* duplicons, the organization of these repeats is remarkably similar. Both sets of repeats are organized in a direct, but non-tandem, fashion, occurring with an average periodicity of once every 22–30 bp (12) (Figs 5 and 7). In addition, degenerate copies of both sequence motifs are common in the repeat tract (Fig. 6). The mechanism by which these sequences promote transposition integration is, as yet, unknown. One possibility may be that these repeats, similarly to the CTGGG repeats of the immunglobulin switch recombination regions (20,28–30), may be hyper-recombinogenic signal sequences which promote the integration of duplicated genomic segments, perhaps, in the form of episomal intermediates (12).

Xq28: a donor hotspot for transposition

Analysis of the duplications of the *ALD* locus defines the second region of Xq28 with extensive autosomal paralogy. The *CTR–CDM* cassette (26.5 kb) demonstrated a similar proclivity to duplicate to various pericentromeric regions in man and higher primates (Fig. 7) (12). Both duplication cassettes show a similar, yet distinct pericentromeric distribution (Figs 1 and 7). These data suggest that this 65 kb portion of Xq28 has been particularly prone to transpose. There are striking similarities between these seemingly independent paralogy domains. Estimates of sequence divergence indicate that both the *ALD* and *CTR–CDM* cassettes duplicated ~5 million years ago (12). Xq28 breakpoints for both duplicons occur near or at inverted clusters of Alu repeats. Both sets of duplications were directed to the pericentromeric regions of the short arm (or long arm of acrocentric) chromosomes (Fig. 7) (12). One possible explanation may be that the entire 65 kb region including the complete copies of the *CTR*, *CDM* and *ALD* genes was duplicated once, possibly to 2p11.1, and that subsequent deletions and transpositions generated two transposon lineages which were duplicated independently to different chromosomes. In this regard, it may be noteworthy that both the *ALD* and *CTR–CDM* duplicons map to the two YACs separated by <500 kb (My895g9 and My662d12) in human 16p11.1–11.2. Alternatively, this particular portion of Xq28 may have been subjected to several independent transposition events which mobilized different segments of Xq28 to various pericentromeric and occasionally coincident locations (Fig. 7).

It has been estimated that 6% of all mutations associated with ALD are the result of sporadic deletions on the X chromosome (1). A comparison of the mapped breakpoints for these deletions (1) with duplication breakpoints (Figs 3 and 7) reveals an interesting association. Of the four deletion patients studied in detail during the molecular cloning of the ALD disease gene (1,31), all share at least one breakpoint within 5 kb of the 5′ transposition breakpoint identified in this study (Figs 3 and 5). The 3′ transposition breakpoint does not show a similar association. These comparisons might suggest that a similar molecular mechanism underlies the proclivity of this region to both delete and transpose. More rigorous examination of the deletion breakpoints at the sequence level is warranted to evaluate this hypothesis critically.

Two-step model for *ALD* transposition

Examination of Xq28–autosome junction sequences among the four *ALD* paralogs reveals that all four sets of breakpoints are identical (Fig. 5). Furthermore, the autosome–autosome paralogy extends in both directions from the Xq28 junctions. This indicates that the mechanism responsible for duplicating and distributing the *ALD* copies in the human genome involves at least two distinct steps. We propose the following model.

Xq28 transposition. A single event mobilized a 9.7 kb Xq28 segment corresponding to the distal portion of the *ALD* locus to a pericentromeric chromosomal region integrating it near a CTTTTTG and Alu monomer repeat junction (Fig. 7). Comparative sequence analysis reveals that *ALD*2p11 has undergone the greatest sequence divergence (Table 2) and, probably, represents the site of the first duplicon integration (>5 million years ago). This is supported by parsimony analysis which indicates that the *ALD*2p11 sequence occupies the deepest branch within the

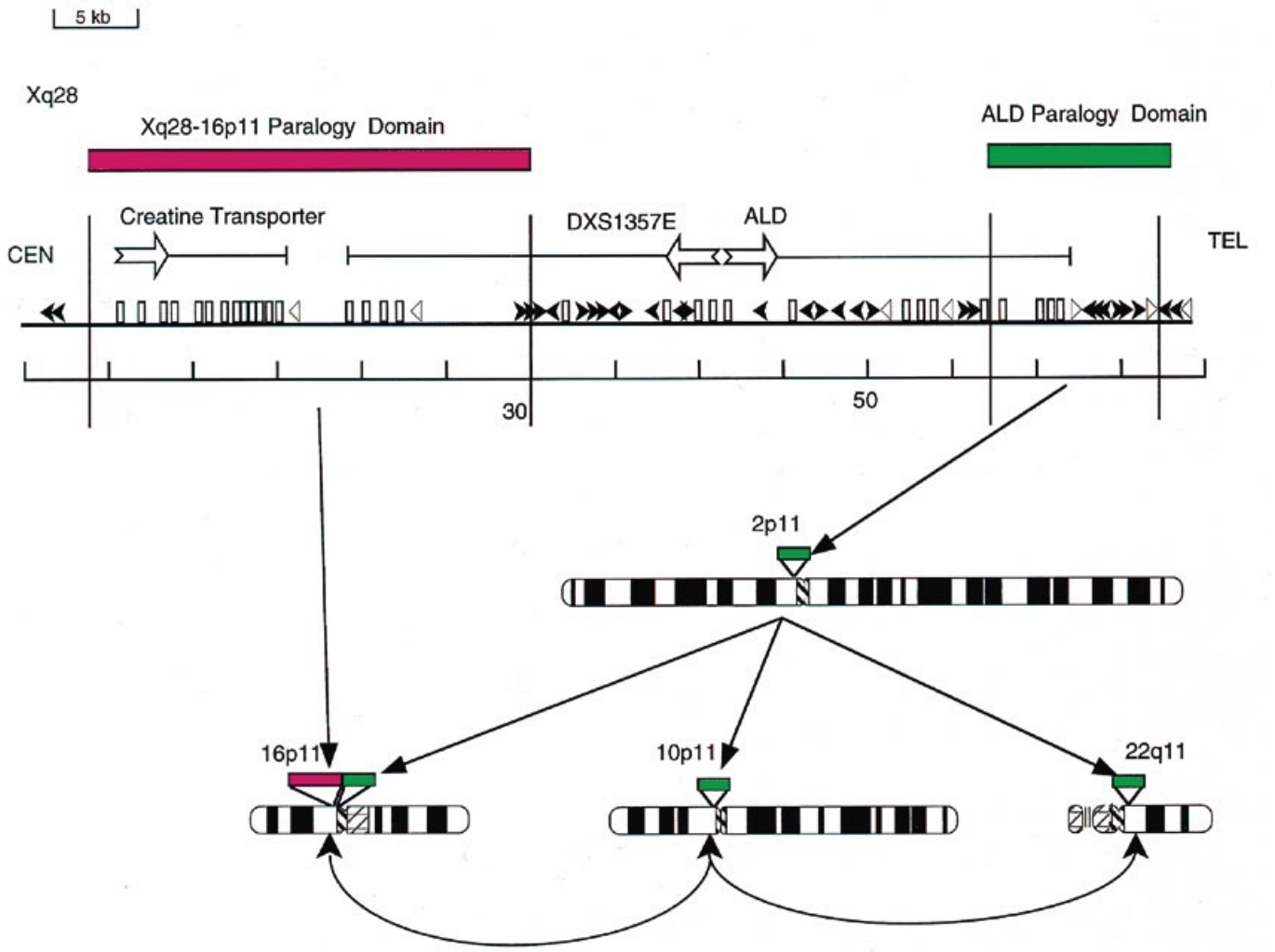

Figure 7. Summary of pericentromeric-directed transpositions. Seventy kilobases of Xq28 sequence spanning the distance from the CTR (*SLC6A8*) to the *ALD* locus are depicted (GenBank U41302 and U52111). The orientation of each gene is indicated by a large arrowhead, and exonic regions are depicted as boxes. Filled arrows indicate the location and orientation of Alu repeats. Other repeat sequences (LINE and MIR) are shown as open arrows. The *CTR–DX1357E* (26.5 kb) duplicon and its transposition is shown by the red horizontal bars. Duplications of the *ALD* cassette (9.7 kb) are shown as a green bar. An initial duplication seeded a copy to 2p11, followed by a transposition burst of this sequence to 16p11, 10p11 and 22q11. The Xq28–autosome breakpoints are shown by vertical lines.

cladogram (Fig. 4). The initial transposition may have involved a larger portion of Xq28 which subsequently may have become deleted (see above).

Pericentromeric exchange. Once the initial ALD autosomal copy had been integrated, it served as a 'seed' for further pericentromeric transposition (Fig. 7). This is supported by the fact that the Xq28–autosome breakpoints are identical and that the paralogy extends in both directions from these junctions (Fig. 6). In addition, parsimony analysis shows that *ALD*2p11, *ALD*22q11 and *ALD*16p11 constitute their own clade, suggesting that the duplicons among the pericentromeric regions are more closely related (Fig. 4). Determining the sequence junctions among the autosome–autosome paralogy domains may shed some light on this second step of pericentromeric exchange. It may be noteworthy, in this regard, that sequencing of the ends of one of the *ALD*16p11.1 cosmids (16–341b10) has identified satellite III sequence homology (data not shown). Centromeric satellite sequences are found frequently to be associated with extrachromosomal polydispersed circular DNA fractions (32,33). It is possible that such episomal intermediates may occasionally include adjacent non-repetitive sequence. If these vectors are capable of integrating into homologous satellite sequence on another chromosome, this could provide the molecular basis for the observed pericentromeric shuttling.

ALD disease implications

The identification and characterization of *ALD* duplicon sequences on four autosomal loci should provide the means for more efficient mutation detection among ALD patients. The extraordinary degree of sequence conservation (Fig. 3) among paralogous exons has probably hampered the identification of *bona fide* mutations associated with ALD. Indeed, several missense mutations commonly reported for ALD patients, such as the G→ A transition at position 2211 (1) and C→T transition

at position 2235 (GenBank Z17859) (5,8,9) are also found among *ALD* paralogous sequences (Fig. 3, positions 65 and 89, respectively). These 'mutations' may represent false-positives detected upon co-amplification of autosomal *ALD* loci. A catalog of duplicated intronic and exonic sequence should facilitate more effective primer design and more critical evaluation of mutations among patients. It should be noted, however, that despite the high degree of conservation (~95%) between autosomal and X-chromosome *ALD* loci, polymorphic autosomal variants from exon 8–10 have rarely been misclassified as *ALD* mutations (3–11).

Evolutionary implications

Our analysis of the duplications of the *ALD* locus to 2p11, 10p11, 16p11 and 22q11 is the third clear-cut example of an unusual phenomenon of pericentromeric-biased transposition (12,22). Extrapolations from earlier data regarding the identification of paralogous gene segments near centromeric regions of chromosomes (17,24,34–37) indicate that the acquisition of genes and gene families within these regions may be a general, hitherto unrecognized property of the pericentromeric portions of primate chromosomes. Analysis of these duplications indicates that most duplicons represent truncated non-processed pseudogenes with little functional significance. Occasionally, however, a gene (such as the *CTR* segment) may be duplicated in its entirety, including its putative promoter (12), maintain transcriptional activity and acquire a new function in an organism (38). Another evolutionary benefit of such pericentromeric plasticity may be to juxtapose different cassettes from diverse genes to create a reservoir of genes in the genome with potentially new functions. Such events, if they did occur, could accelerate an organism's adaptability, allowing for rapid quantitative and qualitative genetic differences within the same species. The evolutionary cost of such pericentromeric plasticity, however, may be that genes of adaptive value occasionally are rearranged or deleted due to their proximity to these areas of 'genetic flux'. It is tempting to speculate that some of the microdeletion syndromes located near the centromeres, such as Prader–Willi/Angelman syndromes in 15q11.2 (39,40), Smith–Magenis syndrome in 17p11.2 (41) and VCF/DG syndromes in 22q11.2 (42,43), may be the consequence of such pericentromeric instability.

MATERIALS AND METHODS

Fluorescent *in situ* hybridization

Chromosome metaphase spreads were prepared from human peripheral blood lymphocytes of a male donor. Cosmid probes were nick-translated, labeled and hybridized to chromosomal preparations as previously described (44). Biotin-labeled cosmid DNA was detected using fluorescein isothiocyanate (FITC)-conjugated avidin (5 µg/ml) (Vector Laboratories). Digoxigenin-11-dUTP (Boehringer Mannheim)-labeled Alu PCR products were co-hybridized to generate an R-banding pattern for cytogenetic band identification (45). A Zeiss Axioskop epifluorescence microscope with a cooled charge-coupled device (CCD) camera was used to generate digital images (Fig. 1).

Library hybridization

Five arrayed chromosome-specific cosmid libraries (LL02NC01 'X', LA10NC02, LA16NC02, LL22NC03 'N' and LLOXNCO1 'U' corresponding to chromosomes 2, 10, 16, 22 and X, respectively), were obtained from Lawrence Livermore and Los Alamos National Laboratories (46,47). A sufficient number of clones were isolated and grown for each library such that >5× coverage was obtained for each chromosome. Filters were pre-hybridized for 1 h at 65°C with 0.25 M $NaPO_4$, 0.25 M NaCl, 5% SDS, 10% polyethylene glycol and 1 mM EDTA; and blocked with 20 µg/ml herring sperm DNA. PCR-generated products (Table 1) were purified (QiaQuick column) and 25 ng of product was random-hexamer labeled (MegaPrime) with [α-^{32}P]dCTP and 1 U of Klenow fragment, according to the manufacturer's specifications (Amersham). Probes were purified through a G-50 Sephadex column and allowed to hybridize overnight at 65°C in a rotisserie oven. Filters were washed three times for 30 min each at 65°C with 0.05 M $NaPO_4$, 0.5% SDS and 1 mM EDTA solution and exposed to autoradiographic film.

Physical mapping

Genomic DNAs (150 ng) from a somatic cell hybrid monochromosomal panel (NIGMS, Human Genetic Mutant Cell Repository) were used as templates in PCR amplification with primer pairs 4 and 11 (Table 1) to confirm the chromosomal location of each *ALD* paralog. Appropriate control DNA from mouse, hamster and human was included in the PCR analysis. To refine the map location for chromosomes 10 and 22, radiation and deletion hybrid panels were analyzed by PCR. For chromosome 10, the following hybrids were tested: 10, 43, 57, 132, 168, 170 and 175 (14). For chromosome 22, hybrids Cl 6-2EG,GM10888, D655, Cl-9/5878 and Cl-4/GB were analyzed (15). In addition, cosmid DNA (22-11c7) containing the chromosome 22 *ALD* paralog was used in FISH as a probe against DiGeorge deletion and translocation patients to map the *ALD* duplicon to the VCF/DG critical region. For chromosome 16, a panel of CEPH YAC clones, 614A5, 653D12, 662D12, 663G12, 693C11, 700E10, 769B3, 897E10, 895G9 and 950B3, which had been STS mapped to the 16p11.1–16p11.2 interval (16) were analyzed by PCR. Cosmid DNA (1 ng) was analyzed by PCR to determine the organization and extent of paralogy (Table 1). With the exception of the annealing temperature (Table 1), PCR conditions were identical (95°C for 3 min; followed by 35 cycles of 95°C for 1 min, 60/65°C for 1 min and 72°C for 2 min). PCR amplification reactions were carried out in a final volume of 25 µl containing 0.25 mM dNTPs (Pharmacia), 25 pmol of each primer and 1.25 U of *Taq* polymerase (Boehringer Mannheim) in standard 1× PCR buffer (Boehringer Mannheim). All cycling conditions were optimized for use in a PE 9600 thermocycler (Perkin Elmer Cetus).

Sequence analysis

PCR products were subcloned both into a TA cloning vector, pGEMT (Promega), using the manufacturer's suggested protocol. Ligation products were transformed into XL1-Blue supercompetent cells (Stratagene), and transformants were screened by PCR to identify clones which contained inserts of the correct length. Positive clones were sequenced with T7 and SP6 primers and fluorescently labeled dideoxy terminators from a single-

strand template using an automated DNA sequencer (ABI 373). Multiple clones from independent ligations were analyzed to confirm the sequence. Sequence analysis of the autosome–Xq28 paralogy breakpoints was performed using the fmol cycle sequencing kit (Promega) and cosmid DNA (2 μg) as template. Primers were developed as close as possible to the breakpoint junction based on Xq28 sequence (GenBank accession no. U52111). Primer 79490, 5′GAAAGCTGGGTGTCCACG-GAGGGAA, was used for analysis of the 5′ breakpoint, and primer 110516, 5′GTACACAGCGACCACTAGGTGAATAC, was used for analysis of the 3′ breakpoint. Primers were end-labeled with [γ-^{33}P]ATP, and sequencing reactions were analyzed on a 6% denaturing polyacrylamide gel.

Phylogenetic analysis

Sequences were aligned using the BestFit program from the GCG software package (Wisconsin Sequence Analysis Package, v. 8). PAUP (phylogenetic analysis using parsimony) version 3.1.1 (Illinois Natural History Survey) was employed to assess phylogenetic relationships based on 790 bp of comparative sequence among the *ALD* paralogs. As gaps in sequence alignments are problematic for phylogenetic analysis, deletions were counted as a single event in these phylogenetic analyses. Parsimony analysis (PAUP) was performed on aligned sequences using exhaustive searches. No ancestral state nor outgroup sequence was defined during the analysis. Two equally parsimonious trees were generated. Bootstrap analysis (100 branch-and-bound replicates) was used to assess the quality of each equally parsimonious tree.

ACKNOWLEDGEMENTS

We are grateful to Dr M. Batzer and E. Nickerson for useful suggestions in the preparation of this manuscript. We would like to thank Drs J.L. Mandel and C.-O. Sarde for providing *ALD* cDNA probes; and J. Tesmer, B. Pesavento, M. Straka, P. Chien and V. Jurecic for technical assistance. This work was performed under the auspices of the U.S. D.O.E. contract No. W-7405-ENG-48 to H.W.M., and supported, in part, by the U.S.D.O.E. contract W-7405-ENG-36 to N.A.D. and L.L.D., and NIH grant DC02027 to M.L.B. Support by Telethon and AIRC to M.R. is acknowledged. E.E.E. is a D.O.E. Distinguished Human Genome post-doctoral fellow.

REFERENCES

1. Mosser, J., Douar, A., Sarde, C., Kioschis, P., Feil, R., Moser, H., Poustka, A., Mandel, J. and Aubourg, P. (1993) Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature*, **361**, 726–730.
2. Sarde, C.-O., Mosser, J., Koschis, P., Kretz, C., Vicaire, S., Aubourg, P., Poustka, A. and Mandel, J.-L. (1994) Genomic organization of the adrenoleukodystrophy gene. *Genomics*, **22**, 13–20.
3. Cartier, N., Sarde, C., Douar, A., Mosser, J., Mandel, J. and Aubourg, P. (1993) Abnormal messenger RNA expresion and a missense mutation in patients with X-linked adrenoleukodystrophy. *Hum. Mol. Genet.*, **2**, 1949–1951.
4. Uchiyama, A., Suzuki, Y., Song, X., Fukao, T., Imamura, A., Tomatsu, S., Shimozawa, N., Kondo, N. and Orii, T. (1994) Identification of a nonsense mutation in ALD protein cDNA from a patient with adrenoleukodystrophy. *Biochem. Biophys. Res. Commun.*, **198**, 632–636.
5. Fanen, P., Guidoux, S., Sarde, C., Mandel, J., Goossens, M. and Aubourg, P. (1994) Identification of mutations in the putative ATP-binding domain of the adrenoleukodystrophy gene. *J. Clin. Invest.*, **94**, 516–520.
6. Feigenbaum, V., Lomard-Platet, G., Guidoux, S., Sarde, C., Mandel, J. and Aubourg, P. (1996) Mutational and protein analysis of patients and heterozygous women with X-linked adrenoleukodystrophy. *Am. J. Hum. Genet.*, **58**, 1135–1144.
7. Fuchs, S., Sarde, C., Wedemann, H., Schwinger, E., Mandel, J. and Gal, A. (1994) Missense mutations are frequent in the gene for X-chromosomal adrenoleukodystrophy (ALD). *Hum. Mol. Genet.*, **3**, 1903–1905.
8. Krasemann, E., Meier, V., Korenke, G., Hunneman, D. and Hanefeld, F. (1996) Identification of mutations in the ALD-gene of 20 families with adrenoleukodystrophy/adrenomyeloneuropathy. *Hum. Genet.*, **97**, 194–197.
9. Ligtenberg, M., Kemp, S., Sarde, C., van Geel, B., Kleijer, W., Barth, P., Mandel, J., van Oost, B. and Bolhuis, P. (1995) Spectrum of mutations in the gene encoding the adrenoleukodystrophy protein. *Am. J. Hum. Genet.*, **56**, 44–50.
10. Braun, A., Kammerer, S., Ambach, H. and Roscher, A. (1996) Characterization of a partial pseudogene homologous to the adrenoleukodystrophy gene and application to mutation detection. *Hum. Mutat.*, **7**, 105–108.
11. Kok, F., Neumann, S., Sarde, C., Zheng, S., Wu, K., Wei, H., Bergin, J., Watkins, P., Gould, S. and Sack, G. (1995) Mutational analysis of patients with X-linked adrenoleukodystrophy. *Hum. Mutat.*, **6**, 104–115.
12. Eichler, E., Lu, F., Shen, Y., Antonacci, R., Jurecic, V., Doggett, N., Moyzis, R., Baldini, A., Gibbs, R. and Nelson, D. (1996) Duplication of a gene-rich cluster between 16p11.1 and Xq28: a novel pericentromeric-directed mechanism for paralogous genome evolution. *Hum. Mol. Genet.*, **5**, 899–912.
13. Neitz, M. and Neitz, J. (1995) Numbers and ratios of visual pigment genes for normal red–green color vision. *Science*, 1013–1016.
14. Moschonas, N., Spurr, N. and Mao, J. (1996) Report of the first international workshop on human chromosome 10 mapping 1995. *Cytogenet. Cell Genet.*, **72**, 99–112.
15. Budarf, M., Eckman, B., Michaud, D., McDonald, T., Gavigan, S., Buetow, K., Tatsumura, Y., Liu, Z., Hilliard, C., Driscoll, D., Goldmuntz, E., Meese, E., Zwarthoff, E., Williams, S., McDermid, H., Dumanski, J., Biegel, J., Bell, C. and Emanuel, B. (1996) Regional localization of over 300 loci on human chromosome 22 using a somatic cell hybrid mapping panel. *Genomics*, **35**, 275–288.
16. Doggett, N., Goodwin, L., Tesmer, J., Meincke, L., Bruce, D., Clark, L., Altherr, M., Ford, A., HC, C., Marrone, B., Longmire, J., Lane, S., Whitmore, S., Lowenstein, M., Sutherland, R., Mudnt, M., Knill, E., Burno, W., Macken, G., Deaven, L., Callen, D. and Moyzis, R. (1995) An integrated physical map of human chromosome 16. *Nature*, **377** (suppl.), 335–365.
17. Borden, P., Jaenichen, R. and Zachau, H. (1990) Structural features of transposed human Vκ genes and implications for the mechanism of their transpositions. *Nucleic Acids Res.*, **18**, 2101–2107.
18. Sarde, C., Thomas, J., Sadoulet, H., Garnier, J. and Mandel, J. (1994) cDNA sequence of Aldgh, the mouse homolog of the X-linked adrenoleukodystrophy gene. *Mamm. Genome*, **5**, 810–813.
19. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) The evolution of genes: the chicken preproinsulin gene. *Cell*, **20**, 555–565.
20. Vogt, P. (1990) Potential genetic functions of tandem repeated DNA sequence blocks in the human genome are based on a highly conserved 'chromatin folding code'. *Hum. Genet.*, **84**, 301–336.
21. Wohr, G., Fink, T. and Assum, G. (1996) A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.*, **6**, 267–279.
22. Regnier, V., Meddeb, M., Lecointre, G., Richard, F., Duverger, A., Nguyen, V., Dutrillaux, B., Bernheim, A. and Danglot, G. (1997) Emergence and scattering of multiple neurofibromatosis (NF1)-related sequences during hominoid evolution suggest a process of pericentromeric interchromosomal transposition. *Hum. Mol. Genet.*, **6**, 9–16.
23. Arnold, N., Wienberg, J., Emert, K. and Zachau, H. (1995) Comparative mapping of DNA probes derived from the Vκ immunoglobulin gene regions on human and great ape chromosomes by fluorescence *in situ* hybridization. *Genomics*, **26**, 147–156.
24. Zimmer, F., Hameister, H., Schek, H. and Zachau, H. (1990) Transposition of human immunoglobulin V kappa genes within the same chromosome and the mechanism of their amplification. *EMBO J.*, **9**, 1535–1542.
25. Tomlinson, I., Cook, G., Carter, N., Elaswarapu, R., Smith, S., Walter, G., Buluwela, L., Rabbitts, T. and Winter, G. (1994) Human immunglobulin VH and D segments on chromosomes 15q11.2 and 16p11.2. *Hum. Mol. Genet.*, **3**, 853–860.
26. Wong, Z., Royle, N. and Jeffreys, A. (1990) A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics*, **7**, 222–234.

27. Archidiacono, N., Antonacci, R., Marzella, R., Finelli, P., Lonoce, A. and Rocchi, M. (1995) Comparative mapping of human alphoid sequences in great apes using fluorescence in situ hybridization. *Genomics*, **25**, 477–484.
28. Hengstschläger, M., Maizels, M. and Leung, H. (1995) Targeting and regulation of immunoglobulin gene somatic hypermutation and isotype switch recombination. *Prog. Nucleic Acids Res. Mol. Biol.*, **50**, 67–99.
29. Arakawa, H., Iwasato, T., Hayashida, H., Shimizu, A., Honjo, T. and Yamagishi, H. (1993) The complete murine immunglobulin class switch region of the alpha heavy chain gene-hierarchic repetitive structure and recombination breakpoints. *J. Biol. Chem.*, **268**, 4651–4655.
30. Davis, M., Kim, S. and Hood, L. (1980) DNA sequences mediating class switching in alpha-immunoglobulins. *Science*, **209**, 1360–1365.
31. Feil, R., Aubourg, P., Mosser, J., Douar, A., Paslier, D., Philippe, C. and Mandel, J. (1991) Adrenoleukodystrophy: a complex chromosomal rearragnement in the Xq28 red/green-color-pigment gene region indicates two possible gene localizations. *Am. J. Hum. Genet.*, **49**, 1361–1371.
32. Assum, G., Fink, T., Steinbeisser, T. and Fisel, K. (1993) Analysis of human extrachromosomal DNA elements originating from different beta-satellite subfamilies. *Hum. Genet.*, **91**, 489–495.
33. Taylor, S., Larin, Z. and Tyler-Smith, C. (1996) Analysis of extrachromosomal structures containing human centromeric alphoid satellite DNA sequences in mouse cells. *Chromosoma*, **105**, 70–81.
34. Frank, S., Klisak, I., Sparkes, R. and Lusis, A. (1989) A gene homologous to plasminogene located on human chromosome 2q11–p11. *Genomics*, **4**, 449–451.
35. Tunnacliffe, A., Liu, L., Moore, J., Leversha, M., Jackson, M., Papi, L., Ferguson-Smith, M., Thiesen, H. and Ponder, B. (1993) Duplicated KOX zinc finger gene clusters flank the centromere of human chromosome 10: evidence for a pericentric inversion during primate evolution. *Nucleic Acids Res.*, **21**, 1409–1417.
36. Tomlinson, I., Cook, G., Carter, N., Elaswarapu, R., Smith, S., Walter, G., Buluwela, L., Rabbitts, T. and Winter, G. (1994) Human immunglobulin VH and D segments on chromosomes 15q11.2 and 16p11.2. *Hum. Mol. Genet.*, **3**, 853–860.
37. Wong, Z., Royle, N. and Jeffreys, A. (1990) A novel human DNA polymorphism resulting from transfer of DNA from chromosome 6 to chromosome 16. *Genomics*, **7**, 222–234.
38. Iyer, G., Krahe, R., Goodwin, L., Doggett, N., Siciliano, M., Funanage, V. and Proujansky, R. (1996) Identification of a testis-expressed creatine transporter gene at 16p11.2 and confirmation of the X-linked locus to Xq28. *Genomics*, **34**, 143–146.
39. Robinson, W., Spiegel, R. and Schinzel, A. (1993) Deletion breakpoints associated with the Prader–Willi and Angelman syndromes (15q11–15q13) are not sites of high homologous recombination. *Hum. Genet.*, **91**, 181–184.
40. Nicholls, R., Fischel-Ghodsian, N. and Higgs, D. (1987) Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. *Cell*, **49**, 369–378.
41. Juyal, R., Figuera, L., Hauge, X., Elsea, S., Lupski, J., Greenberg, F., Baldini, A. and Patel, P. (1996) Molecular analyses of 17p11.2 deletions in 62 Smith–Magenis syndrome patients. *Am. J. Hum. Genet.*, **58**, 998–1007.
42. Halford, H., Lindsay, E., Nayudu, M., Carey, A., Baldini, A. and Scambler, P. (1993) Low-copy-number repeat sequences flank the DiGeorge/velo-cardio-facial syndrome at 22q11. *Hum. Mol. Genet.*, **2**, 191–196.
43. Gong, W., Emanuel, B., Collins, J., Kim, D., Wang, Z., Chen, F., Zhang, G., Roe, B. and Budarf, M. (1996) A transcription map of the DiGeorge and velocardiofacial syndrome minimal region on 22q11. *Hum. Mol. Genet.*, **5**, 789–900.
44. Baldini, A., Miller, D., Shridhar, V., Rocchi, M., Miller, O. and Ward, D. (1991) Comparative mapping of a gorilla-derived alpha satellite DNA on great ape and human chromosomes. *Chromosoma*, **101**, 109–114.
45. Baldini, A. and Ward, D. (1991) *In situ* hybridization banding of human chromosomes with Alu-PCR products: a simultaneous karyotype for gene mapping studies. *Genomics*, **9**, 770–774.
46. Longmire, J., Brown, N., Meincke, L., Campbell, M., Albright, K., Fawcett, J., Campbell, E., Moyzis, R., Hildebrand, C., Evans, G. and Deaven, L. (1993) Construction and characterization of partial digest DNA libraries made from flow-sorted human chromosome 16. *GATA*, **10**, 69–76.
47. Trask, B., Massa, H., Evans, J., Scherer, S., Friedman, C., Youngblom, J., Rouquier, S., Giorgi, D., Martin-Gallardo, A., Wong, D., Iadonato, S., Yokota, H., van den Engh, G., Hearst, J. and Sachs, R. (1995) In Bentley, D., Green, E. and Warterston, R. (eds), *Applications of FISH in Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, p. 14.