

# Complex $\beta$ -Satellite Repeat Structures and the Expansion of the Zinc Finger Gene Cluster in 19p12

Evan E. Eichler,<sup>1,2,3</sup> Susan M. Hoffman,<sup>1,4</sup> Aaron A. Adamson,<sup>1</sup>  
Laurie A. Gordon,<sup>1</sup> Paula McCready,<sup>1</sup> Jane E. Lamerdin,<sup>1</sup>  
and Harvey W. Mohrenweiser<sup>1</sup>

<sup>1</sup>Human Genome Center, BBRP, L-452, Lawrence Livermore National Laboratory, Livermore, California 94550 USA; <sup>2</sup>Department of Genetics, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106 USA

We investigated the organization, architecture, and evolution of the largest cluster (~4 Mb) of *Krüppel*-associated box zinc finger (KRAB-ZNF) genes located in cytogenetic band interval 19p12. A highly integrated physical map (~700 kb) of overlapping cosmid and BAC clones was developed between genetic STS markers D19S454 and D19S269. Using *ZNF91* exon-specific probes to interrogate a detailed *EcoRI* restriction map of the region, ZNF genes were found to be distributed in a head-to-tail fashion throughout the region with an average density of one ZNF duplicon every 150–180 kb of genomic distance. Sequence analysis of 208,967 bp of this region indicated the presence of two putative ZNF genes: one consisting of a novel member of this gene family (*ZNF208*) expressed ubiquitously in all tissues examined and the other representing a nonprocessed pseudogene (*ZNF209*), located 450 kb proximal to *ZNF208*. Large blocks of (~25-kb) inverted  $\beta$ -satellite repeats with a remarkably symmetrical higher order repeat structure were found to bracket the functional ZNF gene. Hybridization analysis using the  $\beta$ -satellite repeat as a probe indicates that  $\beta$ -satellite interspersions between ZNF gene cassettes is a general property for 1.5 Mb of the ZNF gene cluster in 19p12. Both molecular clock data as well as a retroposon-mapping molecular fossil approach indicate that this ZNF cluster arose early during primate evolution (~50 million years ago). We propose an evolutionary model in which heteromorphic pericentromeric repeat structures such as the  $\beta$  satellites have been coopted to accommodate rapid expansion of a large gene family over a short period of evolutionary time.

[The sequence data described in this paper have been submitted to GenBank under accession nos. AC003973 and AC004004.]

Zinc finger (ZNF) genes represent one of the largest gene families in the human genome with an estimated 500–600 members (Hoovers et al. 1992; Becker et al. 1995; Klug and Schwabe 1995). Although the specific function of the majority of ZNF genes remains largely unknown, as a class they are believed to encode transcriptional regulators that in a few instances have been shown to play critical roles in cellular and developmental differentiation processes (Pieler and Bellefroid 1994). ZNF proteins have been implicated in many diverse eukaryotic developmental processes, such as segment pattern formation in the *Drosophila* embryo (Rosenberg et al. 1986); cellular proliferation in the cerebellar

hindbrain of the mouse (Wilkinson et al. 1989); and hematopoietic differentiation among human myeloid precursor cells (Hromas et al. 1991). DNA binding of the encoded proteins is typically mediated by a ZNF motif that consists either of two cysteines and two histidines (*Krüppel* family or  $C_2/H_2$  type) or four cysteines alone (steroid receptor or  $C_2/C_2$  type). The conserved cysteines and/or histidines form a tetrahedral complex around a zinc metal ion, generating a folded loop or “finger” of 30 amino acids that is capable of making contact with DNA (Miller et al. 1985). The number of ZNF motifs is highly variable, especially among the  $C_2/H_2$  type, ranging from 2 to 40 copies in different members of this family (Bellefroid et al. 1991).

The estimated 500 ZNF genes map to a variety of human chromosomes. Fluorescence in situ hybridization (FISH) on metaphase chromosomes us-

<sup>4</sup>Present address: Department of Zoology, Miami University, Oxford, Ohio 45056 USA.

<sup>3</sup>Corresponding author.

E-MAIL eee@po.cwru.edu; FAX (216) 368-3432.

ing various ZNF cDNAs as probes revealed a clustered organization of these genes on chromosomes 7, 9, 10, 12, 16, and 19 (Huebner et al. 1991; Rousseau-Merck et al. 1993; Tommerup and Vissing 1995; Jackson et al. 1996). This clustered organization suggests that tandem duplications were primarily responsible for increasing copy number of this gene family (Thiesen et al. 1991). Chromosome 19 appears to be particularly enriched for ZNF genes with ZNF loci distributed within three gene clusters corresponding to cytogenetic band locations 19p12, q13.2, and q34 (Bellefroid et al. 1993; Shannon et al. 1996). A survey of chromosome 19 C<sub>2</sub>/H<sub>2</sub> type ZNF genes reveals that the majority encode an evolutionarily conserved protein motif of 75 amino acids, termed the *Krüppel*-associated box or KRAB domain (Thiesen et al. 1991; Bellefroid et al. 1993). The KRAB domain has recently been shown to function as a critical domain for protein-protein interaction (Friedman et al. 1996; Kim et al. 1996). Although these genes may theoretically be involved in transcriptional activation or repression, all KRAB-ZNF genes studied to date have been shown to act only as potent transcriptional repressors (Margolin et al. 1994; Witzgall et al. 1994; Pengue et al. 1995; Vissing et al. 1995).

Based on relatively few examples, the exon/intron structure of the KRAB subfamily of C<sub>2</sub>/H<sub>2</sub> type ZNF genes appears to be highly conserved (Bellefroid et al. 1993; Villa et al. 1993; Derry et al. 1995; Baban et al. 1996; Grondin et al. 1996). Members of this gene subfamily characteristically consist of four exons. The first three exons are relatively small and contain the 5' UTR and the KRAB domain, which is split by a single small intron. The last exon contains the spacer region, the tandemly repeated ZNF motif, and the untranslated portion of the KRAB-ZNF gene. Recent reports have suggested the existence of an additional 5' UTR exon consisting of human endogenous retroviral sequences that occasionally appear in alternatively spliced KRAB-ZNF transcripts (Di Cristofano et al. 1995; Baban et al. 1996). The spacer region of the KRAB-ZNF gene is highly variable and is often used to further categorize different members of this subfamily (Bellefroid et al. 1993). There are an estimated 40 closely related KRAB-ZNF finger loci clustered on 19p12 (Bellefroid et al. 1993, 1995), spanning a genomic distance of ~4 Mb. To date, only eight ZNF cDNAs that map generally to the 19p12 band location (*ZNF20*, *ZNF43*, *ZNF85*, *ZNF90*, *ZNF91*, *ZNF92*, *ZNF93*, and *ZNF94*) have been identified and characterized (Bellefroid et al. 1991, 1993; Lichter et al. 1992). Northern blot analysis with a few 19p12

cDNAs indicates low levels of transcription in multiple tissues, including undifferentiated myeloid cells as well as embryonal carcinoma cell lines (Bellefroid et al. 1993).

Southern and Northern "zooblot" comparisons using 19p12 ZNF spacer-encoding probes detect no cross-hybridization signals in either mouse or hamster (Bellefroid et al. 1993, 1995). These observations have been confirmed by FISH analysis, using *ZNF91* YAC pools as probes, on chromosomal metaphase spreads from a variety of primate species, which indicates that *ZNF91* synteny in this region of chromosome 19 does not extend to prosimians and rodents (Bellefroid et al. 1995). These data suggest that the emergence and expansion of the 19p12 cluster of KRAB-ZNF genes have occurred relatively recently during primate evolution [~64 million years ago (mya)] (Bellefroid et al. 1995). This has led to the supposition that the function of the *ZNF91* gene family cannot be involved in fundamental developmental processes common to the anthropoid and prosimian primate lineages, rather that their function may be related more to chromatin modulation as opposed to gene-specific transcriptional regulation (Bellefroid et al. 1995). The recent identification of a KRAB corepressor protein (KAP-1 or KRIP-1) that contains protein domains (PHD, bromo, and RING finger motifs) common to eukaryotic chromatin-modulating genes gives some support to this hypothesis (Friedman et al. 1996; Kim et al. 1996). There are relatively few examples of recent evolutionary expansions of gene families in the primate lineage (Teglund et al. 1994; Eichler et al. 1996, 1997; Regnier et al. 1997; Zimonjic et al. 1997). It is likely that specific molecular mechanisms exist for the propagation and expansion of selectively favored gene families in any genome. To investigate more directly the molecular basis of the recent expansion of the *ZNF91* gene cluster in 19p12, a detailed BAC/cosmid physical map was constructed in a 700-kb interval of this cluster and large-scale sequence analysis was performed within two selected regions. Our analysis provides the first detailed insight into the genomic architecture of the KRAB-ZNF gene family and implicates higher order  $\beta$ -satellite repeat structures in the expansion of this gene family in the anthropoid genome.

## RESULTS

### Construction of a (700-kb) Physical Map between D19S454 and D19S269

We developed a highly integrated physical map of

cosmid and BAC clones between markers D19S454 and D19S269 (Mohrenweiser et al. 1996) using a combined approach of FISH, STS markers, and conversion of overlapping sets of YAC genomic clones to cosmid and later BAC clones within this interval (Fig. 1). Four overlapping YAC clones (784E8, 411H1, 138D1, and 60D10) spanning an estimated 1.75 Mb of this region were initially used as probes to screen a flow-sorted chromosome 19 arrayed cosmid library. A total of 85 cosmids were identified, assigned to bins, and assembled into overlapping sets of cosmid clones based on fluorescent fingerprinting methods (Carrano et al. 1989; Trask et al. 1993). The location and relative position of each cosmid contig were confirmed using an STS screening strategy and FISH analysis of sperm pronuclei with representative cosmid clones from each contig as probes (Brandriff et al. 1991, 1994) (Fig. 1b). Five sets of contiguous cosmid clones were initially identified within the D19S454/D19S269 interval (Fig. 1c), representing ~550 kb of the 700-kb interval (Fig. 1). Using existing chromosome 19 cosmid clones as walking probes, larger insert BAC clones that bridged adjacent but nonoverlapping contigs were identified (see Methods; Fig. 1). Comparative *EcoRI* digests and subsequent BAC to cosmid hybridizations confirmed the position and orientation of each BAC clone within the region. Because the BAC and chromosome 19-specific cosmid libraries originate from two different genomic sources, comparative *EcoRI* digests between these two sources served as an important check and balance in confirming the integrity of the genomic clones used in the construction of a physical map of this region. Two BAC/cosmid contigs were generated in this region (representing 300 and 500 kb of chromosome 19 sequence), separated by a gap of ~25 kb (Fig. 1c), as determined by distance estimates using high-resolution two-color FISH.

### ZNF Orientation and Duplicon Size within 19p12

Owing to the reported highly conserved nature of ZNF genes in 19p12 (Bellefroid et al. 1991, 1993), exon-specific PCR products corresponding to the KRABA, KRABB, spacer, and ZNF domains of *ZNF91* were developed and used as probes to screen nylon-transferred *EcoRI* digests of portions of the D19S454/S269 interval (Fig. 1c). All probes hybridized strongly to paralogous *EcoRI* fragments within this region with the exception of the KRABB probe, which demonstrated variable degrees of hybridization signal intensity (data not shown). Using this cross-hybridization approach, it was possible to de-

termine the position of ZNF genes in the region, their orientation, and the distance separating paralogous ZNF genes. Within the D19S454/S269 region we identified at least four genes that are organized in a head-to-tail fashion. The direction of putative transcription of each gene is telomeric to centromeric. The average duplicon size for each gene in this region was ~180 kb (Fig. 1c). In comparison, a second region was analyzed that is located ~1 Mb proximal to the D19S454/S269 region, abutting the  $\alpha$ -satellite repeat region of the chromosome 19 centromere. Here, the distance between two adjacent ZNF genes was found to be smaller (~100 kb), suggesting variable compaction of ZNF genes across the 4-Mb cluster (data not shown).

### Comparative Sequence Analysis

Two genomic clones within the D19S454/S269 region were selected for large-scale sequence analysis: BAC 33152 (~150 kb in length) and cosmid 32532 (~40 kb in length). Based on the physical map and earlier hybridization experiments, each clone contained a putative ZNF gene. These genes were separated by a distance of ~450 kb (Fig. 1c). Random shotgun M13 libraries were prepared for each clone, subclones were sequenced, and sequence data were assembled using PHRAP software. The actual inserts of BAC 33152 and cosmid 32532 were determined to be 165,199 bp and 43,768 bp, respectively, for a total of 208,967 bp of 19p12 ZNF genomic sequence. The finished sequences for BAC 33152 and cosmid 32532 have been deposited in GenBank under accession numbers AC003973 and AC004004, respectively.

An overview of the comparative genomic organization and the ZNF gene structures of BAC 33152 and cosmid 32532 is presented below (Table 1; Fig. 2). A combination of GRAIL analysis (v. 1.3) and BLAST sequence similarity comparisons against cDNA sequence from ZNF gene family members was used to determine the most likely position of intron-exon boundaries within each clone. Cosmid 32532, owing to the length of its insert, contained only three (KRABA, KRABB, and spacer-ZNF exons) of the four exons (the fourth exon, predominantly 5' UTR, was not present in this clone). BAC 33152 contained a complete complement of ZNF exons. The two putative ZNF genes showed the greatest nucleotide sequence similarity (~83%) with two previously identified members of the ZNF 19p12 gene family, *ZNF43* and *ZNF91*, and were designated *ZNF208* (BAC 33152) and *ZNF209* (cosmid 32532) in accordance with the GDB nomenclature commit-

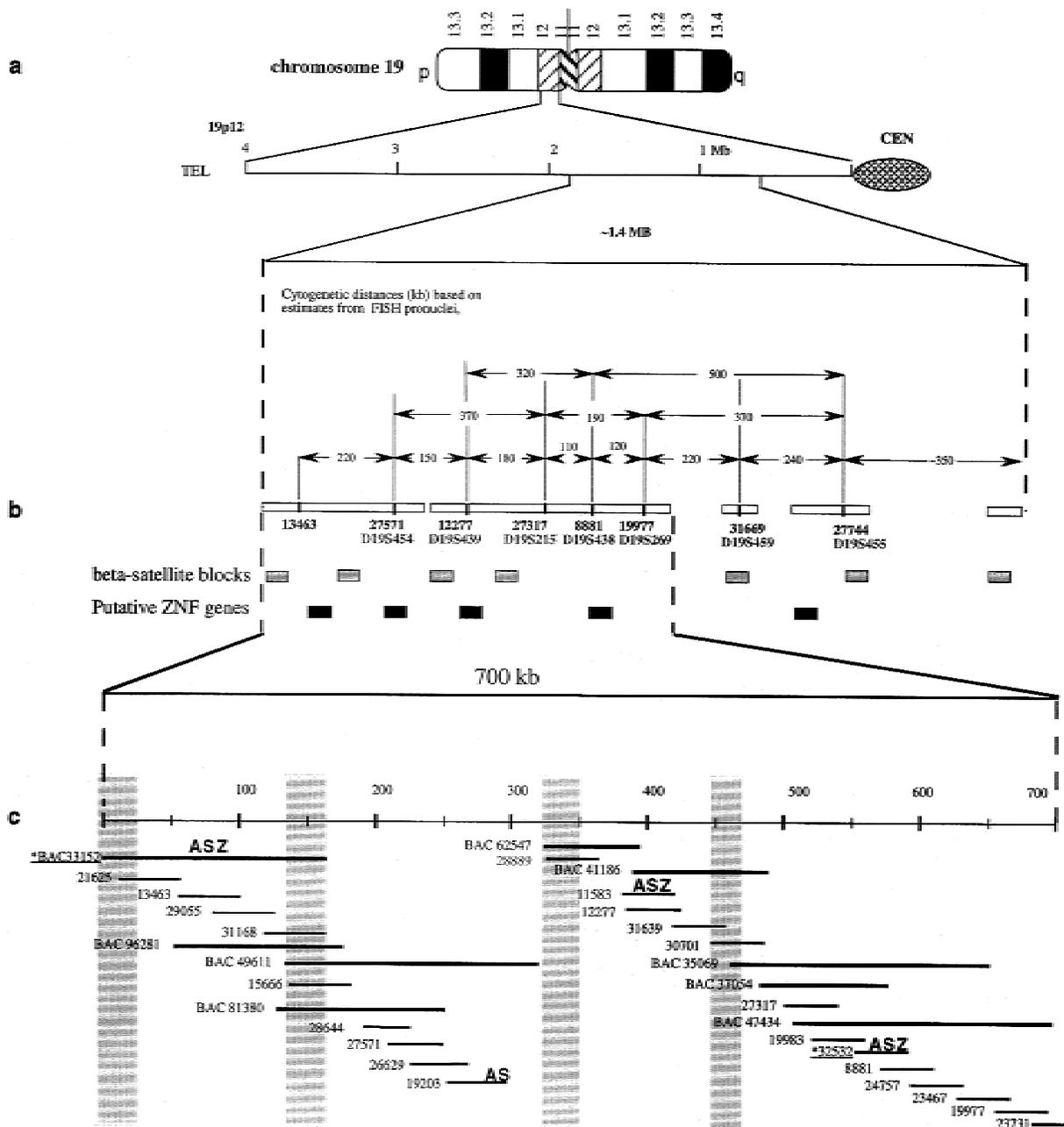


Figure 1 Physical map of the ZNF gene cluster in 19p12. The organization of the region under study is depicted at three levels of resolution. (a) An ideogram of chromosome 19 delineates the ~1.5-Mb portion from the 4- to 5-Mb region of the ZNF cluster that has been characterized. (b) Cytogenetic distances between representative cosmids are shown based on estimates of physical distance from three-color FISH analysis on decondensed chromatin from sperm pronuclei. Cosmids used in the construction of the framework cyto-genetic map are indicated below each contig (shown as open horizontal bars) in association with genetic markers for the chromosome 19 map. (c) Overlapping BAC and cosmid genomic clones for 700 kb of this region are shown as horizontal lines. The orientation and order of clones in the contig were determined based on an *EcoRI* fingerprinting strategy of multiple overlapping clones. Only a subset of the clones in the total tiling path of these clones is indicated. The positions of cross-hybridizing probes specific for the KRABA exon (A), spacer region (S) and zinc finger gene (Z) of *ZNF91* are indicated within the physical map. Vertical shaded bars indicate the position and extent of regions that cross-hybridize to  $\beta$ -satellite probes derived from BAC 33152. The two clones that have been sequenced are indicated with an asterisk (\*).

Table 1. Gene Structure

Gene	Exon	Position	Length (bp)	Intron (bp)
<i>ZNF208</i>	1 (5' UTR)	42654–42772	118	21,883
<i>ZNF208</i>	2 (KRABA)	64655–64781	126	1,472
<i>ZNF208</i>	3a (KRABB)	66253–66348	95	12,409
<i>ZNF208</i>	3b (KRABB + 3' UTR)	66253–66818	565	—
<i>ZNF208</i>	4 (ZNF + 3' UTR)	78757–83719	4962	—
<i>ZNF209</i>	1 (5' UTR)	?	?	?
<i>ZNF209</i>	2 (KRABA)	254–380	126	497
<i>ZNF209</i>	3 (KRABB)	878–973	95	6,314
<i>ZNF209</i>	4 (ZNF + 3' UTR)	7287–10096	2809	—

The exon/intron structures of gene *ZNF208* and pseudogene *ZNF209* are shown. Exon boundaries were determined based on cDNA comparisons of other *ZNF91* family members to the genomic sequence and by GRAIL analysis. Positions of exons refer to GenBank accession no. HUAC003987 for *ZNF208* and GenBank accession no. HUAC004004 for *ZNF209*.

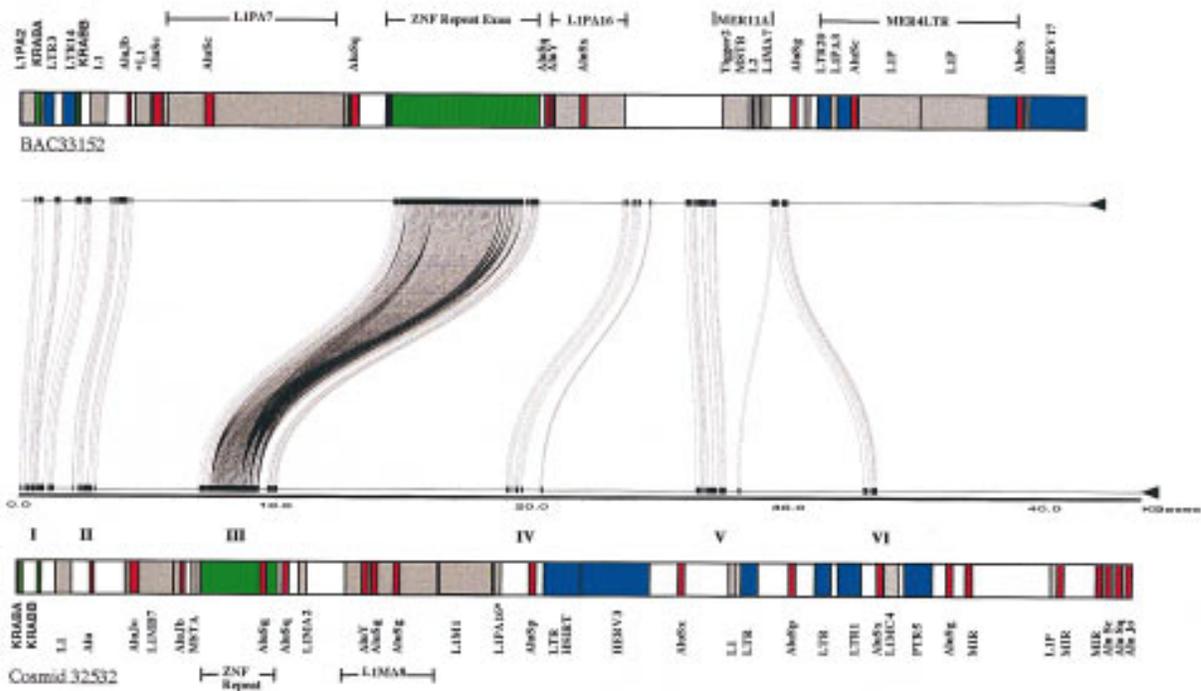


Figure 2 Comparative sequence analysis of ZNF duplicons. A genomic segment, corresponding to the entire insert (43,378 bp) of cosmid clone 32532 (GenBank accession no. HUAC003973) and paralogous positions 64,401–108,079 from BAC 33152 (GenBank accession no. HUAC004004), was compared between the two ZNF 19p12 clones using Miropeats software (threshold score  $s = 25$ ; setting = onlyinter). Before analysis, common repeat elements were removed from the sequence using RepeatMasker software. Regions of sequence conservation that do not carry LINE or SINE sequences are delineated by “joining” lines between the two sequences. Six regions of paralogy were identified (indicated by Roman numerals), and each region was aligned using the BESTFIT alignment program (GCG software package). The position and identity of exons and other repeat elements are illustrated schematically above the miropeat alignment. Similar results were obtained using dot-matrix alignment software with unmasked sequence. Note that the identity and position of most repeat elements are not conserved between the duplicated segments.

tee. The gene structures of *ZNF208* and *ZNF209* are generally conserved and are similar to the *ZNF91* model (Bellefroid et al. 1993), consisting of separate exons for KRABA and KRABB protein binding domains and a single large exon that incorporates the ZNF DNA-binding domain and spacer region. The sizes of the KRABA and KRABB exons (exons 2 and 3) are highly conserved between *ZNF208* and *ZNF209* and are predicted to be both 126 and 95 bp in length, respectively (Table 1). With the exception of exons 2 and 3, the gene structure of *ZNF208* is more compact than *ZNF209*. The size of the last putative exon, for example, differs significantly. Exon 4 of *ZNF209* contains 15 identifiable ZNF (28-amino-acid) repeats over 2.8 kb, whereas the same exon from *ZNF208* (from BAC 33152) is almost twice as large, 4.9 kb, with 41 ZNF repeat motifs. Similarly, the lengths of introns 2 and 3 are almost twice as long in *ZNF208* as in *ZNF209*.

Using Miropeats, RepeatMasker, and DOTTER software (Parsons et al. 1993; Sonnhammer and Durbin 1995), we compared the nongenic organization of BAC 33152 and cosmid 32532. Two 43-kb segments that corresponded to the complete insert of cosmid 32532 (43,768 bp, GenBank accession no. AC004004) and the paralogous segment of BAC 33152 (64401–108,079 bp, GenBank accession no. AC003973) were masked for low-copy repeat sequences, and the two sequences were compared using miropeats (Fig. 2). This analysis identified six regions (~10.5 kb) of relatively high sequence similarity (77.4%–82.0%) between the two duplicated segments of 19p12 (Fig. 2). Not surprisingly, three of these regions corresponded to the positions of exon/introns of the putative ZNF genes. The other three regions, however, were located distal to the final exon and were not associated with any known genic or repeat sequences. An analysis of LINE and SINE retroposons between these two segments found very little conservation in the organization and subfamily identity of these repeat elements in the region, suggesting that the majority of Alu and L1 retroposition invasions occurred independently within the genomic context of these two ZNF cassettes. Only a single, fragmentary L1 element (726 bp in length located in intron 3) showed complete conservation in orientation, position, and identity between the two sequences. Sequence analysis of this LINE element shows that it belongs to a relatively ancient subfamily (L1MB), which was predominantly active during the time of the mammalian radiation. BESTFIT (GCG software) alignment of noncoding paralogous segments between BAC 33152 and cosmid 32532 revealed 79.4% sequence

identity over a compared region of 8041 bp (conserved segments I, II, IV, V, and VI; Fig. 2). This degree of sequence similarity, however, does not persist throughout the entire 43-kb segment, largely owing to the differential organization of L1 and Alu repeat elements between the two regions. Sequence similarity (83.1%) was observed between the putative coding portions of *ZNF208* and *ZNF209*.

### Expression Analysis

BLAST sequence similarity searches with the putative ZNF genic portion of BAC 33152 against the National Center for Biotechnology Information (NCBI) dbEST database identified a single EST from a human pregnant uterus cDNA library (IMAGE clone 501492) with 99.9% identity over 500 bp. Subsequent complete sequence analysis of the cDNA insert (809 bp) showed nearly 100% sequence identity between the cDNA clone and the genomic sequence of BAC 33152. The cDNA sequence extends from the 5' UTR region (exon 1) through the KRABA and KRABB portions of *ZNF208* (exons 2 and 3), terminating 470 bp distal to the transcriptional splice donor of exon 3a near an alternative poly(A) addition signal (AATATA) (exon 3b; Table 1). Translation of the ORF of EST 501492 predicts a 77-amino-acid protein, consisting only of KRABA and KRABB protein-interacting domains. Conceptual translation of exon 4 from the BAC clone, however, suggests an additional 3.7 kb of ZNF motifs that is not part of cDNA clone 501492. Three strong polyadenylation signals were identified near the terminal portion of the ZNF repeat motifs predicting a transcript of ~5.0 kb. Based on this analysis, two different ORFs are expected for *ZNF208*: a short peptide (77 amino acids) completely devoid of ZNF repeat motifs and a longer isoform (ORF = 3951 bp/1317 amino acids) consisting of 41 ZNF (28-amino-acid) repeats.

Northern blot analysis of 16 human tissues using EST 501492 PCR-amplified insert did not show strong signal hybridization to a single transcript. Instead, a high level of background (multiple weak hybridization signals of ~1.0 kb and 4.2, 4.4, and 5.0 kb) was observed in most tissues examined. Such weak background hybridization signals have been reported previously for several ZNF genes (Derry et al. 1995; Baban et al. 1996; Ogawa et al. 1998) and, as has been suggested, likely represent a combination of low-level expression and cross-hybridization from multiple members of the KRAB-ZNF gene family. To eliminate the problem of background hybridization and to specifically test for each of the two

*ZNF208* isoforms, a specific RT-PCR assay was developed for each of the two alternative transcripts of the *ZNF208* gene (Fig. 3). RT-PCR products consistent with the two different mRNA isoforms for *ZNF208* were identified in most tissues examined (Fig. 3). Sequence analysis of the RT-PCR products confirmed that these amplification products repre-

sented expression from *ZNF208*. These data confirm that *ZNF208* is a bona fide gene with two different transcripts resulting from alternative splicing and utilization of different polyadenylation signal sequences.

Database searches with the putative ZNF genic portion of cosmid 32532 identified no ESTs with

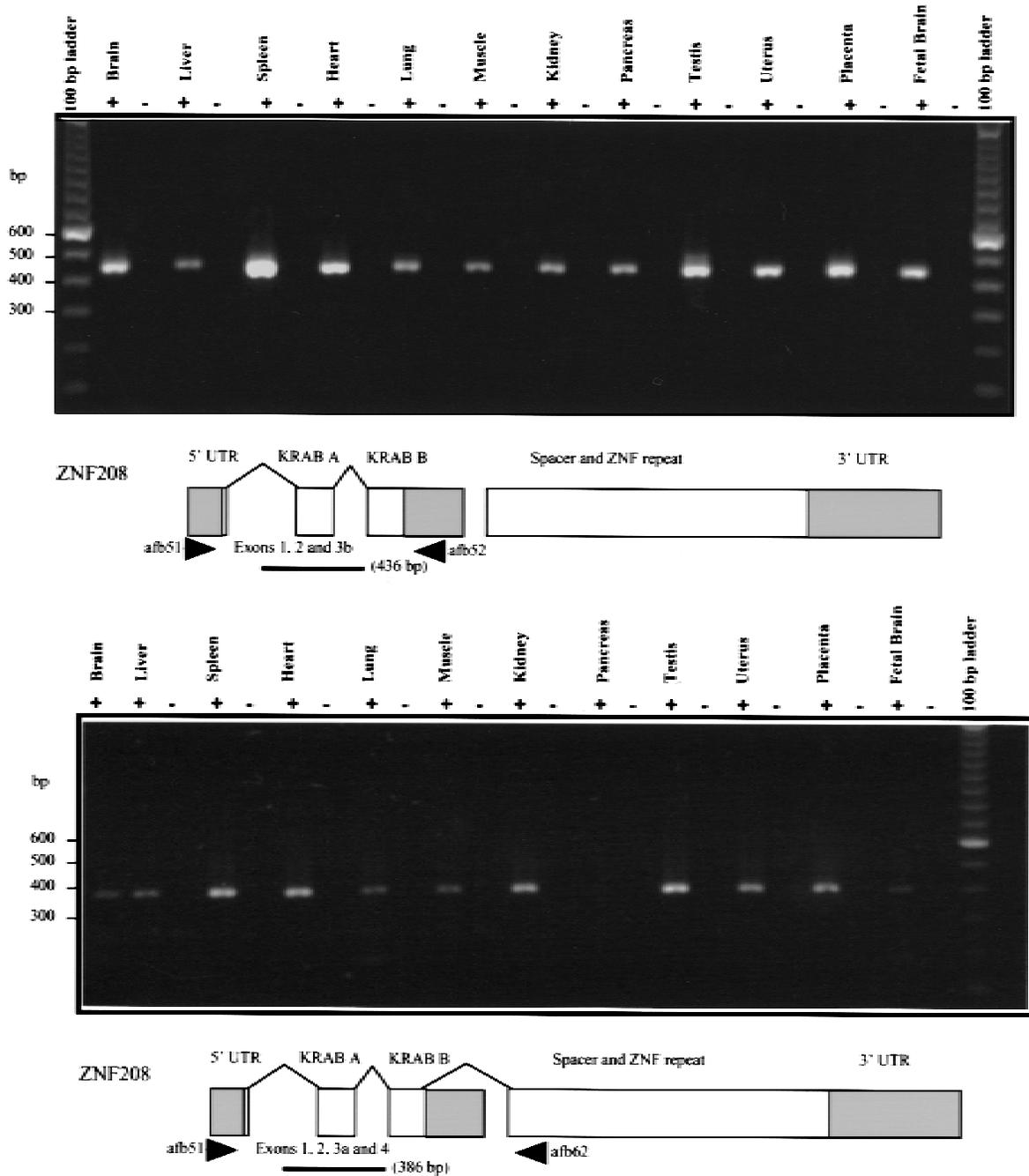


Figure 3 RT-PCR analysis of *ZNF208*. RT-PCR analysis of *ZNF208* from 12 human tissue sources confirmed the presence of two transcripts that result from the usage of alternative polyadenylation signals. The position of the primers and concomitant PCR product size are shown with respect to the *ZNF208* gene structure. cDNA synthesis reactions without reverse transcriptase for each tissue served as a negative control (indicated by a minus sign).

sequence similarity greater than the background level of homology for the *ZNF91* gene family cluster (85%–90%). Conceptual translation of *ZNF209* predicted multiple stops (~6 stop codons) in-frame with normal ZNF translation. RT-PCR using two different primer pairs failed to detect expression of *ZNF209* in any of the tissues examined. These data argue that *ZNF209* is a nonprocessed pseudogene.

### $\beta$ -Satellite Repeat Structures Flanking 19p12 ZNF Genes

Examination of the genomic organization of BAC 33152 using Miropeats, RepeatMasker, and dot-matrix analysis software revealed the presence of large blocks (~24–45 kb in size) of inverted repeat structures flanking the *ZNF208* transcription unit (Fig. 4). Sequence similarity searches indicated that large portions of these repeat structures showed lim-

ited homology (75%–84%) to previously characterized  $\beta$ -satellite consensus motifs (GenBank accession no. M81228) (Greig and Willard 1992). In addition to inverted  $\beta$ -satellite structures, human endogenous retrovirus sequences were also observed in opposite orientation flanking the ZNF gene within BAC 33152. These observations suggest an inverted symmetry to the organization of repetitive elements in the vicinity of *ZNF208*.

Analysis of the  $\beta$ -satellite blocks flanking *ZNF208* revealed a remarkable higher order superstructure. Three  $\beta$ -satellite superstructures ranging in length from 21 to 23 kb were defined within BAC 33152 (Fig. 5; Table 2). Each of these structures was found to consist of three portions: (1) a 5'  $\beta$ -satellite segment of 5–10 kb, (2) a middle Alu/LTR portion of 2–3 kb, and (3) a 3'  $\beta$ -satellite segment of 5–10 kb. Two of these are located adjacent to one another, suggesting that they may compound to form a larger 40-kb structure (Fig. 4).

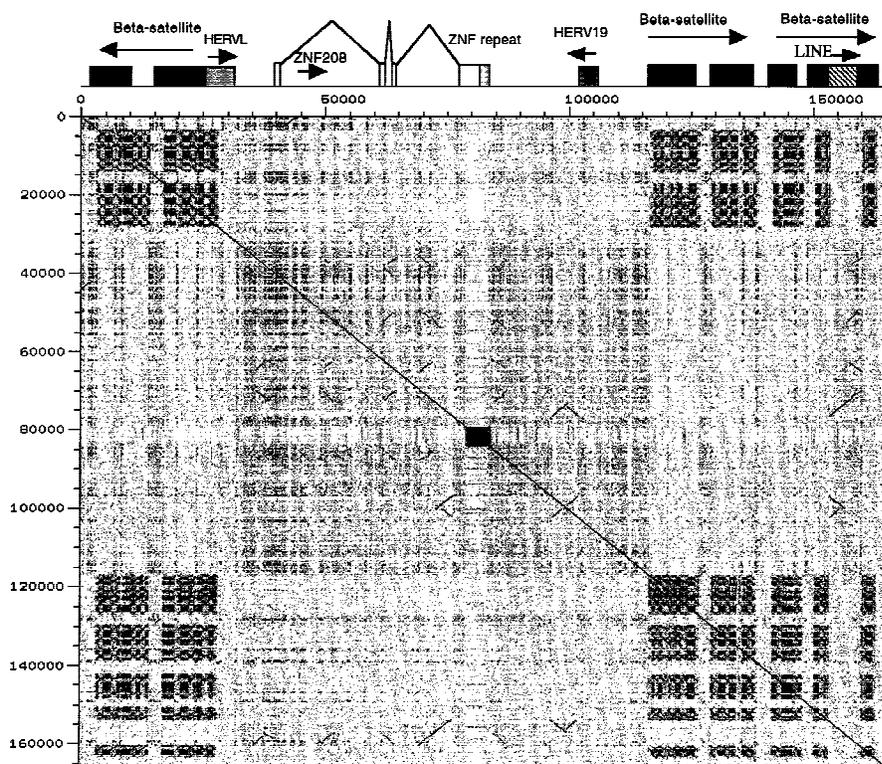


Figure 4 Genomic organization of BAC 33152. A schematic diagram depicting the general organization of BAC 33152, including the positions of exons as well as the human endogeneous retroviral elements and  $\beta$ -satellite repeat regions (~20–25 kb in length) that flank the *ZNF208* gene. The  $\beta$ -satellite repeat sequences flanking the gene are inverted in orientation with respect to one another. The genomic organization is placed in the context of a dot-matrix alignment of BAC 33152 sequence (AC004004) against itself (DOTTER). The ZNF repeats appear as a black square symmetrically located in the center, whereas the three  $\beta$ -satellite repeat superstructures appear as a "patchwork crosses" on either side of the ZNF gene.

The marked symmetry in both length and orientation of repeat elements is shown below (Fig. 5). If interspersed repeat elements (LINEs and SINEs) are excluded from the calculation, the length of the 5' and 3'  $\beta$ -satellite portions with respect to the LTR/Alu complex segment appear highly conserved for each  $\beta$ -satellite superstructure (Fig. 5). Interestingly, in all three  $\beta$ -satellite blocks examined, an Alu repeat element was observed symmetrically located within the center of the  $\beta$ -satellite repeat element. In two of the three blocks examined, the Alu repeat element is conspicuously oriented in an inverted orientation with respect to other repeat elements within the superstructure.

During the analysis of the  $\beta$ -satellite repeat structures, it was found that the individual  $\beta$ -satellite repeat units were not simply organized as tandem reiterations of the 68-bp consensus motif within each  $\beta$ -satellite segment (Agresti et al. 1989; Waye and Willard 1989; Willard 1990). Instead,

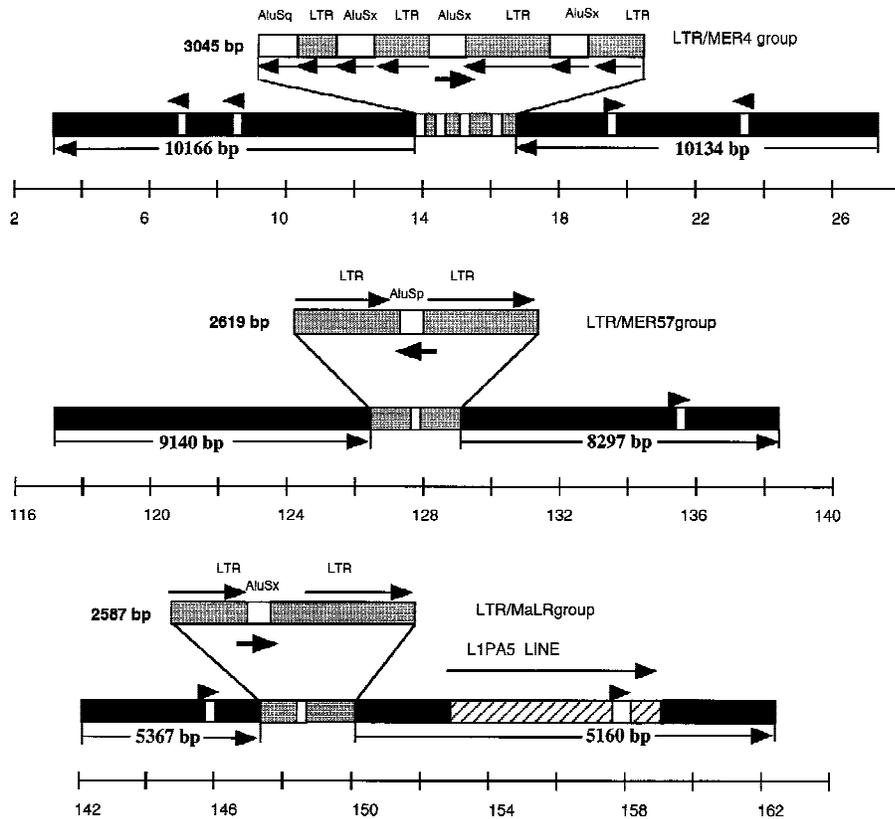


Figure 5 Higher order structure of  $\beta$ -satellite repeats.  $\beta$ -satellites are organized into super-repeat structures consisting of two  $\beta$ -satellite segments flanking an Alu/LTR middle portion. The organization of each of the three "superstructures" is drawn to scale with reference to the positions of the repeats in GenBank (accession no. AC004004). The total length of each of the three segments in bp is indicated beside each segment. Calculation of length of the  $\beta$ -satellite flanks did not include Alu and LINE elements. This is especially evident for the structure located at positions 140–160, in which an L1PA5 element has integrated.

the repeats are distributed as clusters of tandem arrays of variable length with intervening sequence separating each cluster. A total of 322  $\beta$ -satellite repeat sequences were found to be distributed in 26 clusters comprising 26,573 bp of the total sequence of BAC 33152. Using MEME (multiple expectation-maximization for motif elicitation) software (Bailey and Elkan 1994), a 71-mer consensus motif was generated from these 322 repeats (Table 2). The most favored consensus motif demonstrates 78.9% sequence identity with the previously identified " $\beta$ -satellite" consensus sequence (Agresti et al. 1989; Wayne and Willard 1989; Willard 1990) (Fig. 6). Sequence similarity searches with each intervening sequence between each cluster of  $\beta$ -satellite repeats showed no significant homology to known repeat sequences in the NCBI database. Dot-matrix analysis of these regions, however, indicated that different intervening segments exhibit low-level se-

quence similarity to each other. This suggests that the intervening segments themselves are repetitive. MEME analysis further revealed that each intervening segment is composed of tandem reiterations of a degenerate 38-mer repeat motif (A total of 389 repeat units were identified within 22 clusters embedded within 21,036 bp of sequence). The most favored consensus motif of this 38-mer repeat shows 65.9% sequence similarity to a core segment of the  $\beta$ -satellite consensus motif (Fig. 6). Thus, the  $\beta$ -satellite segments are composed of alternating clusters of repeats demonstrating high sequence similarity (the 71-mer repeat) and low sequence similarity (the 38-mer repeat) to the  $\beta$ -satellite consensus motif.

To investigate whether these  $\beta$ -satellite repeat structures were a general property of the architecture of the ZNF cluster in 19p12, a 1.5-kb  $\beta$ -satellite probe (M13 clone afb69e9, corresponding to positions 142121–153021 of GenBank accession no.

AC003973) was hybridized against an arrayed (10 $\times$  coverage) chromosome 19 cosmid library. A total of 95 chromosome 19-specific cosmids were identified that hybridized intensely with the  $\beta$ -satellite repeat probe. Fifty-two of these cosmids were distributed among five contigs whose location within 19p12 had been determined previously using high-resolution two-color FISH and STS content mapping (see Methods). Analysis of the locations of the  $\beta$ -satellite structures within the *EcoRI* by Southern hybridization, as well as inference of positive/negative hybridizing cosmids within each contig, predicts that there are at least seven blocks of  $\beta$  satellites (~40 kb in size), spanning 1.5 Mb of the 4-Mb cluster of ZNF genes. Comparison of the positions of the  $\beta$ -satellite blocks with the position of putative ZNF genes indicates that these repeat structures generally bracket ZNF genes (see Fig. 1). These large blocks of  $\beta$  satellites occur with a periodicity of once every

Table 2.  $\beta$ -Satellite Substructure of BAC 33152

Repeat	Position	Size	Number	orientation
71mer	3159-3705	546	6	-
35mer	3706-4684	978	22	-
71mer	4685-6267	1582	21	-
35mer	6268-7583	815	20	-
71mer	7584-10493	2909	32	-
35mer	10494-11514	1020	22	-
71mer	11515-12402	887	9	-
35mer	12403-13348	947	22	-
71mer	13349-13736	387	5	-
LTR/Alu complex	13843-16888	3045	NA	-
71mer	16889-19309	2420	30	-
35mer	19310-20644	730	20	-
71mer	20645-21401	756	9	-
35mer	21402-21805	403	5	-
71mer	21806-22138	332	4	-
35mer	22139-23221	1082	22	-
71mer	23222-24725	1205	15	-
35mer	23945-26884	1958	34	-
71mer	26685-27243	738	4	-
ZNF/HERV region	27872-117324	89452	NA	-
71mer	117332-117731	399	4	+
38mer	117732-18861	1129	22	+
71mer	118862-119333	471	6	+
38mer	119334-122254	2677	31	+
71mer	122255-123033	778	8	+
38mer	123034-124433	1398	29	+
71mer	124433-125945	1512	19	+
38mer	125946-126670	724	12	+
LTR/Alu complex	126671-129290	2619	NA	-
71mer	129321-129879	558	6	+
38mer	129880-130940	1060	22	+
71mer	130941-132330	1389	17	+
38mer	132331-133181	850	18	+
71mer	133182-134420	1238	16	+
38mer	134421-135851	710	17	+
71mer	135852-136824	972	12	+
38mer	136825-137750	927	19	+
71mer	137751-138346	595	8	+
IS	138381-141914	3533	NA	-
71mer	141929-142914	985	8	+
38mer	142915-143909	899	21	+
71mer	143910-146735	2524	25	+
38mer	146736-147160	424	8	+
71mer	147161-147696	535	6	+
LTR/Alu complex	147942-150529	2587	NA	-
71mer	150544-151644	1100	12	+
38mer	151645-152539	895	17	+
71mer	152540-152898	358	4	+
LINE	153474-160168	6694	NA	-
71mer	160206-160441	238	3	+
38mer	160442-161261	819	16	+
71mer	161262-162421	1159	11	+
38mer	162422-163012	591	12	+
LTR	163013-164461	1448	NA	-

The  $\beta$ -satellite repeat segments consist of alternating clusters of a 71-bp repeat motif and a 38-bp repeat motif, as determined by MEME analysis. The length of each cluster, the number of repetitive units, and their orientation with respect to the  $\beta$ -satellite consensus sequence are summarized. Alu and LINE elements were excluded when estimating the size of each cluster of repetitive motifs.

150–200 kb, a pattern of reiteration consistent with the tandem duplication of the ZNF genes (Fig. 1). The predicted “beads-on-a-string” architecture of  $\beta$ -satellite repeats in this region of 19p12 was subsequently confirmed by FISH analysis using  $\beta$ -satellite repeats as probes against alkaline-borate preparations of metaphase nuclei (data not shown).

### DISCUSSION

We have investigated the genomic organization of the ZNF gene cluster located in cytogenetic band interval 19p12 at several levels of scrutiny. To obtain a general overview of the ZNF gene cluster organization, we constructed an integrated physical map (~700 kb) of overlapping cosmid and BAC clones between genetic markers D19S454 and D19S269. Based on hybridization experiments with *ZNF91* exon-specific probes, we identified four potential ZNF genes arranged in a head-to-tail fashion in this region with an average periodicity of one ZNF gene every 150 kb. This ZNF gene density in 19p12 is in general agreement with earlier estimates of the size of this gene cluster that predicted ~40 different genes within the 4- to 5-Mb interval of 19p12 (Bellefroid et al. 1991). A recent study into the organization of KRAB-ZNF genes in a different ZNF cluster located in 19q13.2 found a much greater density of genes, with as many as 15 different ZNF genes duplicated over a distance of 350–450 kb (Shannon et al. 1996). Interestingly, the average size of the 19q13.2 ZNF duplicon is more than five times smaller than the spacing between ZNF genes in 19p12. The discovery of large (25- to 40-kb)  $\beta$ -satellite structures located on either side of some ZNF genes in 19p12 indicates that these repeat elements may account for some of the differences in spacing between these clusters. Similar to the 19p12 cluster, however, 19q13.2 ZNF genes were found to be arranged in a head-to-tail fashion. This suggests that both ZNF gene clusters have most likely arisen by a common evolutionary mechanism involving

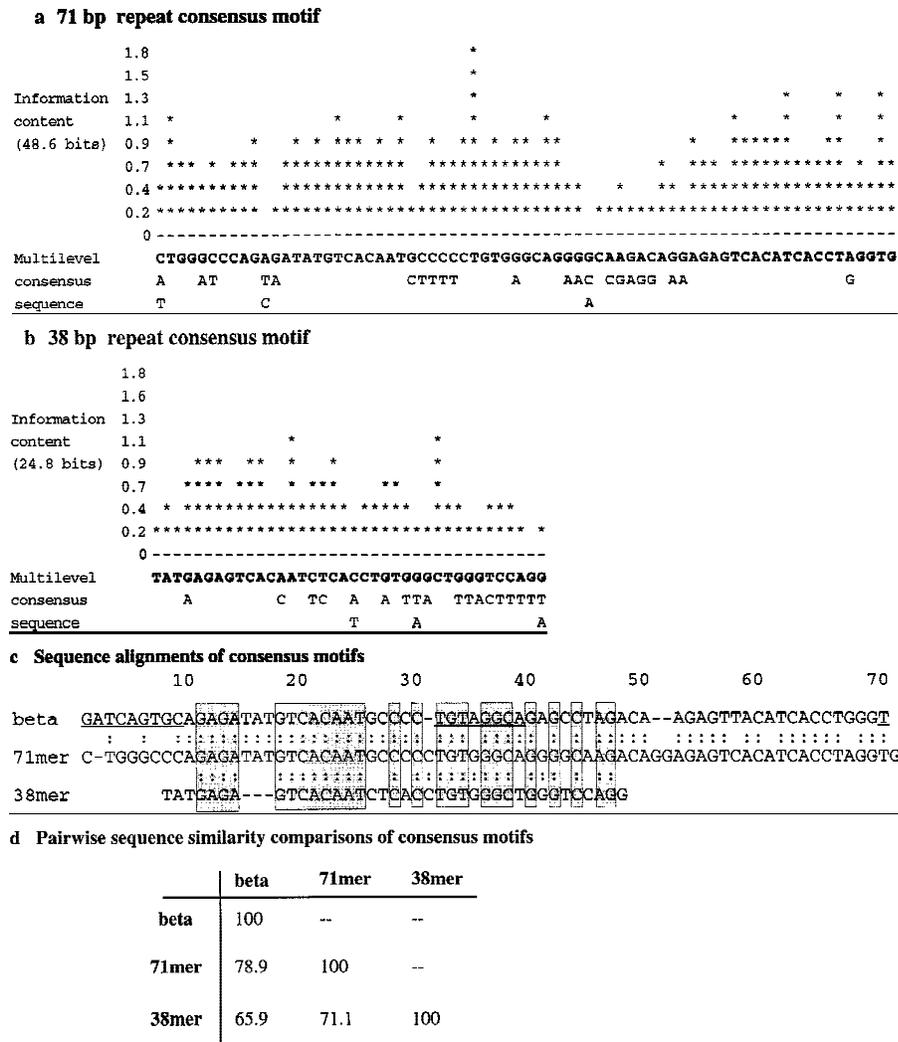


Figure 6 Beta-satellite consensus motifs. (a) A multilevel consensus sequence was generated using MEME software analysis of 322 motifs of 71 bp in length over 26,573 bp of BAC 33152 sequence. These regions were analyzed together based on sequence similarity to β-satellite consensus sequence. The information content (described in bits) provides a relative measure of the degree of conservation for each basepair position in the consensus motif. The most favored consensus is shown in boldface with less favored bases shown below each position. A given basepair is only included in the multilevel consensus if it occurs with a frequency of >0.2 in the consensus. (b) A multilevel consensus sequence was similarly constructed based on MEME analysis of 389 motifs of 35 bp in length over 21,036 bp. This analysis was performed on those sequences that were located between regions showing sequence similarity to β-satellites. (c) BESTFIT alignment of the most-favored consensus motifs for the 38-mer and 71-mer repeats against the β-satellite consensus (Vogt 1990) is shown. Sequence that is conserved among all three repeat elements is shaded and boxed. Underlined sequence indicates regions highly conserved among previously identified β-satellite repeat units. (d) Percentage pairwise sequence similarity (BESTFIT software, GCG) is shown for the three consensus motifs.

subsequent divergence of individual family members.

The general model for the structural organization of 19p12 ZNF genes consists of four exons: a 5' UTR exon that includes the translational initiation codon, two exons encoding KRABA and KRABB protein interacting domains, and a fourth exon that contains the spacer region, the DNA-binding ZNF repeat domain and the 3' UTR (Bellefroid et al. 1991). Our analysis of 208,967 bp of 19p12 identified two potential ZNF genes (Bellefroid et al. 1991, 1993) whose intron/exon structures were in complete agreement with the *ZNF91* model (Table 1). The fact that a nearly identical exon/intron structure has been observed in other chromosome 19 ZNF clusters as well as KRAB-ZNF genes from other chromosomal locations (Villa et al. 1993; Derry et al. 1995; Constantinou-Deltas et al. 1996; Grondin et al. 1996) suggests that this modular organization is a general property of most KRAB-ZNF genes in the human genome. Conceptual translation, cDNA sequencing, and RT-PCR expression analysis indicated that *ZNF208* is a functional gene that is expressed in most tissues (Fig. 2). Two distinct splice variants were identified for *ZNF208*, one of which is comprised of only the KRABA and KRABB protein domains and results from the utilization of an alternative suboptimal polyadenylation signal (AATATA) before splicing of the fourth exon. Although the functional significance of these ZNF "tailless" transcripts remains to be determined, one hypothesis is that KRABA and KRABB peptides devoid of their

endoduplication of an ancestral "seed" ZNF cassette to generate a tandem array of genes followed by

subsequent divergence of individual family members.

DNA/RNA-binding motif function to sequester proteins that normally interact with full-length ZNF proteins to corepress transcription of target genes (Baban et al. 1996; Friedman et al. 1996; Kim et al. 1996). Such a competition for interaction with KRAB corepressors such as KAP-1 may prevent the association of these repression protein complexes with DNA/RNA.

To evaluate the evolutionary age of the expansion of the ZNF cluster, we compared 43 kb of sequence from two different nonadjacent ZNF gene cassettes from the 19p12 cluster (Fig. 3). Sequence conservation was identified by Miropeat and dot-matrix analysis in six genomic regions (totalling ~10.2 kb) between the *ZNF208* and *ZNF209* duplicons. BESTFIT alignment of the noncoding portions of these duplicons (8041 bp) showed 79.5% nucleotide identity. Based on the neutral mutation rate ( $5 \times 10^{-9}$  mutations per site per year), we estimate that the two duplicated segments diverged from a common ZNF ancestral sequence ~40 mya. The remaining ~32.8 kb, which showed virtually no sequence homology, consisted almost entirely of different short interspersed repeat elements. It may be noteworthy that the majority of retroposons identified in these two duplicated segments belong to subfamilies that were active before the divergence of the Old World and New World monkeys (35–44 mya) (Shen et al. 1991; Smit 1993; Smit and Riggs 1995; Smit et al. 1995; Batzer et al. 1996). These include the L1PA7, L1PA16, HERV3, and AluS retroelements (Fig. 3). The fact that these occur in nonparalogous positions between *ZNF208* and *ZNF209* suggests that the duplicated ZNF structure already existed before the divergence of these anthropoid clades. Although more detailed analysis of other *ZNF91* duplicated genomic segments from 19p12 is required, both the molecular clock and molecular fossil data would indicate that the expansion of the *ZNF91* gene cluster in 19p12 occurred ~40–50 mya. These findings are in general agreement with other studies that showed no association of *ZNF91* genes with the syntenic region of 19p12 among prosimians, such as tarsier and squirrel monkey, but identified 19p12 orthologous *ZNF91* genes in all anthropoids studied (Bellefroid et al. 1991).

Sequence analysis of the intergenic regions of *ZNF208* uncovered a complex genomic architecture of  $\beta$ -satellite repeats (Fig. 4) bracketing this ZNF gene. Three hierarchical levels of organization were identified. First, large blocks (25–40 kb) of  $\beta$ -satellite repeat sequences flanking a core segment harboring a ZNF gene appear to define the basic unit of ZNF duplication for a significant portion of the 19p12

gene cluster. Second, peculiar superrepeat structures were identified ranging in length from 20 to 23 kb within each “block” of  $\beta$  satellites. Each of the three  $\beta$ -satellite superstructures of BAC 33152 showed a similar tripartite organization: a 5'  $\beta$ -satellite repeat portion, a middle portion consisting of a complex of Alu and LTR retroelements, and a 3'  $\beta$ -satellite repeat portion (Fig. 5). Remarkable symmetry was observed for each of these structures in which the Alu/LTR complex was centrally located flanked by  $\beta$ -satellite “arms” of nearly identical length (Fig. 5; Table 2). Such a conservation in length of genomic segments harboring  $\beta$  satellites is particularly surprising because the number of  $\beta$ -satellite repeats, especially among acrocentric chromosomes, is known to be unstable and highly variable within the human population (Willard 1990). Finally, the basic chromosome 19  $\beta$ -satellite units (71 bp) were organized into clusters of tandem repeats ranging from 3 to 34 monomers (Table 2). These clusters were separated by intervening segments (~800 bp in length) that were themselves a 38-mer degenerate repeat of the  $\beta$ -satellite repeats (Fig. 6).

Previous studies into the organization of  $\beta$ -satellite repeats have not indicated such a complicated higher order repeat organization (Worton et al. 1988; Willard 1990). Most investigations have suggested that  $\beta$  satellites are organized as large tandem arrays ranging in length from 70 to 400 kb abutting other satellite DNAs in the vicinity of the centromere (Cooper et al. 1992; Shiels et al. 1997). Although the chromosome 19 organization may be exceptional, large-scale sequence analysis will be required to determine whether similar higher order repeat structures are present in other  $\beta$ -satellite chromosomal regions. It is intriguing that large palindromic structures have been identified recently for a different pericentromeric repeat multisequence family, termed chAB4 (Assum et al. 1991; Wohr et al. 1996). chAB4 repeat units are organized as inverted duplications of 90 kb flanking a “nonduplicated” core sequence estimated to be ~60 kb in length. This is similar to the ZNF- $\beta$  satellite organization observed in this study in which a 40-kb inverted duplication of  $\beta$ -satellite repeats flanks a 90-kb core sequence harboring the ZNF gene. Interestingly, both of these palindromic structures appear to be localized exclusively within the pericentromeric regions of chromosomes and are organized as clusters within each of these regions. As has been suggested, such inverted structures may be an inherent property in the dispersal and proliferation of these repeat sequences (Wohr et al. 1996).

In the human genome,  $\beta$ -satellite repeats have

been identified in the pericentromeric regions of chromosomes 1, 9, and Y, as well as all acrocentric chromosomes (13, 14, 15, 21, and 22) (Agresti et al. 1987, 1989; Waye and Willard 1989; Willard 1990; Cooper et al. 1992; Greig and Willard 1992). To our knowledge, this is the first report describing the presence of β-satellite repeats in 19p12. Several features of the chromosome 19 β-satellite repeats, however, are anomalous with respect to the classically defined β-satellite (68-mer) repeats of other chromosomes. The chromosome 19 β-satellite repeats reiterate with a periodicity of once every 71 bp instead of 68 bp, and they lack the highly conserved nucleotide block (GATCAGTGC) that has been proposed to function as a protein-binding site for this repeat (Agresti et al. 1989; Vogt 1990). Although the overall consensus motif exhibits 78.9% identity to the "standard" β-satellite repeat consensus, of the 322 repeat units examined in BAC 33152, not a single repeat showed >75% identity to previously characterized β-satellite repeat units. The interposed 38-mer β satellite-like clusters between the 71-mer repeats showed substantially less sequence similarity (Fig. 6). It is not surprising, then, that previous fluorescent in situ experiments with other β-satellite probes failed to identify the presence of β-satellites on 19p12. Reciprocal experiments with probes derived from the β-satellite repeat structures of BAC 33152 hybridized exclusively to 19p12 (data not shown) indicating that these particular repeats are specific to chromosome 19. This suggests that other large satellite repeat sequences distributed in the pericentromeric region of the human genome may yet remain to be discovered.

Both pericentromeric and telomeric regions of human chromosomes have recently been shown to demonstrate an unusual proclivity to duplicate gene-containing genomic segments (Eichler et al. 1996, 1997; Winokur et al. 1996; Regnier et al. 1997; Zimonjic et al. 1997; Trask et al. 1998). It has been suggested that various repeat sequences in these regions may be involved in promoting duplication. The fact that many of the KRAB-ZNF genes are located in close proximity to subtelomeric and pericentromeric regions may explain their rapid proliferation in the human genome (Lichter et al. 1992; Tommerup et al. 1993; Tunnacliffe et al. 1993; Tommerup and Vissing 1995; Hoffman et al. 1996; Jackson et al. 1996). Because β satellites appear to have duplicated in concert with the ZNF genes in this region of 19p12, the most prosaic explanation is that they were part of the original ZNF gene cassette that became duplicated. We propose a model in which a single ancestral ZNF progenitor gene be-

came associated with β-satellite repeat sequences located in proximal 19p12 (Fig. 7). This may have occurred by a process of pericentromeric-directed transposition as has been described for other human chromosomes or by chromosomal rearrangements that are common during speciation events. In this regard, it is interesting that FISH experiments with *ZNF91* cDNA probes against prosimian chromosomal metaphase spreads have identified putative orthologs in subtelomeric regions of chromosomes that are not syntenic to 19p12 (Bellefroid et al. 1995). Centromeric repeat sequences, such as β-satellites, are known to be capable of rapid expansion and contraction, presumably by mechanisms involving saltatory replication or unequal crossing-over events (Willard 1990). Once the ancestral ZNF gene integrated near such a heteromorphic β-satellite repeat, it began to be duplicated, becoming effectively carried within the β-satellite matrix that was in a state of flux. The presence of inverted β-

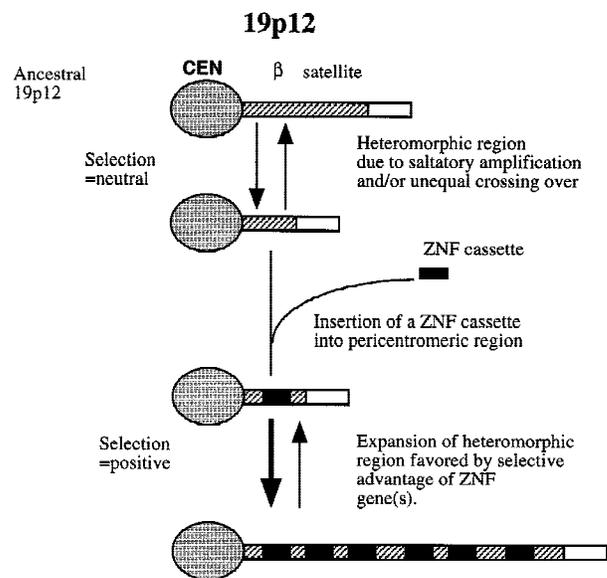


Figure 7 Model for the expansion of the ZNF gene cluster in 19p12. A hypothetical model is proposed in which the pericentromeric region of 19p12 is in a state of expansion and contraction owing to saltatory amplification and/or unequal crossing-over of β-satellite repeats in this region. A functional ZNF progenitor gene associates with β-satellite repeats in 19p12 ~50 mya. The region continues to expand and contract, effectively carrying the inserted ZNF gene as part of its heteromorphism. Expansion becomes favored over contraction among the β-satellites owing to the placement of a functional gene within its context that confers a selective advantage. This leads to the formation of a large cluster of tandemly duplicated ZNF genes in the anthropoid ancestor.

satellite blocks flanking the ZNF gene may have promoted further duplication events. Such large palindromes have been shown to promote gene amplification somatically (Hyrien et al. 1988; Windle and Wahl 1992) and are found associated with other duplicated gene family clusters (Bishop et al. 1985; Groot et al. 1990; Gao et al. 1997). Owing to the potential selective advantage of expressed ZNF genes within the repeat structure, evolutionary pressure may have favored expansion over contraction of this region, leading to the generation of a large cluster of tandemly duplicated genes in the human genome (Fig. 7). Such a model, although speculative, would help explain the rapid expansion of the *ZNF91* gene family over a relatively short period of time during primate evolution (Bellefroid et al. 1995). It should be emphasized that an association between ZNF genes and  $\beta$  satellites has only been documented for a 1.5-Mb region of 19p12 located in proximity to the centromere. It will be interesting to determine whether  $\beta$  satellites or perhaps other pericentromeric repeat sequences have been involved in the expansion of the remaining ~3.0 Mb of the ZNF gene cluster in this region.

## METHODS

### Physical Map Construction

A foundation physical map of sets of overlapping cosmids between STS markers D19S454 and D19S269 was constructed as described previously (Ashworth et al. 1995). More than 85 cosmid clones were identified from chromosome 19-specific libraries [LLN19C02 "F," LLN19C03 "R" (de Jong et al. 1989)] that hybridized to YAC clone probes from this region. The organization of the framework cosmid map was confirmed by fluorescence in situ hybridization in sperm pronuclei to estimate distance between selected cosmid clones in the map (Fig. 1), the assignment of known chromosome 19 STS markers to the region, an automated fluorescence-based restriction fingerprinting technique to confirm the order and overlap of *EcoRI* fragments, and hybridization of known YAC insert probes [Centre d'Etude du Polymorphisme Humain (CEPH) YAC library] that had been assigned to the region (Ashworth et al. 1995). A human genomic BAC library (5 $\times$  coverage) (Research Genetics) was screened to identify larger insert clones that would bridge adjacent but nonoverlapping cosmid contigs. A previously described protocol involving long-range inter-Alu PCR in conjunction with T7-Alu and Sp6-Alu PCR (Parrish et al. 1995) was used to generate probes from terminal cosmids of each set of overlapping clones. Hybridization against total human genomic BAC libraries identified a set of candidate BAC clones that were each, in turn, subjected to long-range inter-Alu PCR amplification, and the fragments were used as probes back against the chromosome 19-specific cosmid libraries. This cosmid-to-BAC and BAC-to-cosmid approach was used to ensure the isolation of bona fide chromosome 19 BACs and to identify additional cosmid contigs that mapped to the region. Comparative *EcoRI* fluorescent

fingerprinting between cosmid contigs and BACs as well as FISH hybridization on chromosomal preparations from human metaphase and sperm pronuclei (Trask et al. 1993) were used as final criteria for assigning BACs to the 19p12 physical map.

### Library Preparation and Sequencing

Random shotgun libraries from BAC 33152 and cosmid 32532 were prepared in the M13mp18 vector using slight modifications of a previously described protocol (Lamerdin et al. 1996). Cosmid and BAC DNA was isolated using Quiawell 8 DNA Isolation Systems (Qiagen), and the DNA was sheared using a TDL Nebulizer (constructed at the Washington University School of Medicine) for 4 min at 30 psi to generate DNA fragments with an average insert size of 1.5 kb. Size-selected and end-repaired fragments were blunt-end ligated and subcloned into the M13mp18 vector. Well-separated M13 plaques were arrayed into 96-well microtiter plates using an automated colony picker designed by Lawrence Berkeley Laboratory. After incubation at 37°C for 7–8 hr, single-stranded DNA for each subclone was isolated using either Qiagen 96-well format M13 kits (per manufacturer's specifications) or a previously described PEG precipitation protocol (Kristensen et al. 1987). Fluorescent dye-primer sequence reactions were prepared using a Catalyst 800 Molecular Biology labstation, a PE9600 Thermocycler, and ABD *Taq* thermosequencing cycle sequencing kits (Perkin-Elmer Applied Biosystems). A modified asymmetric PCR protocol (Munzy et al. 1993) was used in the directed reverse sequencing phase of the project. Sequencing reaction products were analyzed on both ABD 373A and ABD 377 sequencers (Applied Biosystems). The sequence data was analyzed using PHRED/PHRAP software, and the assembled sets of overlapping sequence reads were edited using Consed v.3.0 (software available from Phil Green and David Gordon; <http://genome.wustl.edu>). Regions lacking double-stranded continuity or areas of poor sequence quality within the sequence assembly were identified by SWEDISH software (available from Matt Nolan, Lawrence Livermore National Laboratory). Additional subclones and sequence reads were generated within these regions. One region (2.2 kb) was directly subcloned from BAC 33152, and a sublibrary, using the ABI PRISM Primer Island Transposition Kit (PE Applied Biosystems), was prepared to increase the number of high-quality double-stranded sequencing reads in the region (Devine and Boeke 1994). A total of 590 sequence reads were analyzed and assembled for the 43-kb insert of cosmid 32532, (11.7-fold sequence redundancy, 99.5% double stranded). A total of 3419 sequence reads were analyzed and assembled for the 163-kb insert of BAC 33152 (10.7-fold sequence redundancy, 98.5% double stranded). The sequence of BAC 33152 and cosmid 32532 was deposited in GenBank under the accession numbers AC003937 and AC004004.

### RT-PCR Analysis

Alternative splicing of *ZNF208* was analyzed by RT-PCR using three oligonucleotide amplification primers: afb51 (5'-TCCTACTGCTGTGTGCCTCTGCTCC-3'), afb52 (5'-CTACTTCTTTGGAAACACAGCTTCCAG-3'), and afb62 (5'-TTCTATGCCCTGCTCTGGCCAAAG-3'). High stringency PCR conditions were optimized to eliminate cross-

amplification from other ZNF loci and ensure *ZNF208* specificity. afb51/afb52 amplification conditions consisted of an initial denaturation of 5 min at 95°C, followed by 35 cycles of 30 sec at 95°C, 30 sec at 65°C, and a 45-sec extension at 72°C. A final extension of 5 min was carried out at 72°C. afb51/afb62 RT-PCR reactions were similar, with the exception that both extension and annealing profiles were combined into a single amplification step of 45 sec at 75°C. All cycling conditions were performed in a 9600 Thermocycler (Perkin-Elmer Applied Biosystems). The cDNA was prepared with the Superscript cDNA synthesis kit following manufacturer's suggested protocol (GIBCO BRL). Poly(A) mRNA (1 µg; Clontech) isolated from 12 human tissue sources (adult brain, liver, spleen, heart, lung, muscle, kidney, pancreas, testis, uterus, placenta, and fetal brain) was used as template in cDNA preparation. For each cDNA synthesis reaction, a negative control without the reverse transcriptase was included (Fig. 3). PCR products from brain, testis, and uterus were cloned for each RT-PCR reaction using pGEMT-Easy (Promega), and plasmid DNA was isolated (Qiagen) and sequenced using standard dye-primer fluorescent chemistry (Applied Biosystems). A minimum of three clones were sequenced for each RT-PCR reaction.

### Computer Software

The location and identity of various repeat elements in the sequences were determined using RepeatMasker software (<http://ftp/genome/washington.edu/cgi-bin/RepeatMasker>). Miropeat software (Parsons 1995) was used to identify other blocks of internal repeat sequences that were not contained within the RepMask database. DOTTER, a dot-matrix sequence alignment program (Sonnhammer and Durbin 1995), in conjunction with Miropeats determined the extent of paralogy between duplicated segments of BAC 33152 and cosmid 32532. In addition, both software programs were used to determine the general architecture of the β-satellite repeat motifs flanking *ZNF208* (Fig. 4). The locations of putative exons in the sequences were determined with the GRAIL-2 gene-recognition tool (Uberbacher and Mural 1991) and by comparisons of known *ZNF91* cDNA sequences to the genomic sequence. All sequence alignments were performed using BESTFIT software (GCG). A GeneWorks software package (v. 2.1, Intelligenetics) provided conceptual translation of putative coding regions in the sequences. To identify repeat consensus motifs within the β-satellite repeat regions, the software tool MEME (Bailey and Elkan 1994) was used. Regions showing sequence similarity to the β-satellite consensus (>70%–80%) were extracted from BAC 33152 (total basepairs = 26,573) and subjected to MEME analysis. A similar analysis was performed on intervening repeat-masked sequences (total basepairs = 21,306) located between regions of β-satellite homology. The most probable consensus motifs were generated from comparison of 322 sequence motifs for the 71-mer repeat and 389 sequence motifs for the 38-mer repeat. MAST software (motif alignment search tool) was used to evaluate the significance of each consensus motif (Bailey and Elkan 1994).

### ACKNOWLEDGMENTS

We thank A.F. Smit and T.L. Bailey for assistance in the analysis of repeats in this region and M. Coleman for helpful suggestions in the preparation of this manuscript. We are grateful

to M. Christensen, L. Woo, A. Kobayashi, and D. Ow for excellent technical and informatics assistance. The chromosome-specific gene libraries LL019NC02 and LL019NC03 used in this work were constructed at the Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA 94550. This research was supported, in part, by a U.S. Department of Energy (DOE) Human Genome Postdoctoral Fellowship to E.E.E., by a Howard Hughes Medical Institute grant to Case Western Reserve University School of Medicine, and by the Lawrence Livermore National Laboratory under the auspices of DOE contract W-7045-Eng.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

### REFERENCES

- Agresti, A., G. Rainaldi, A. Lobbiani, I. Magnani, R. Di Lernia, R. Menerver, A. Siccardi, and E. Ginelli. 1987. Chromosomal location by in situ hybridization of the human Sau3A family of DNA repeats. *Hum. Genet.* 75: 326–332.
- Agresti, A., R. Menerver, A. Siccardi, A. Marozzi, G. Corneio, S. Gudi, and E. Ginelli. 1989. Linkage in human heterochromatin between highly divergent Sau3A repeats and a new family of repeated DNA sequence (HaeIII family). *J. Mol. Biol.* 205: 625–631.
- Ashworth, L., M. Batzer, B. Brandriff, E. Branscomb, P. de Jong, E. Garcia, J. Ganes, L. Gordon, J. Lamerdin, G. Lennon et al. 1995. An integrated metric physical map of human chromosome 19. *Nature Genet.* 11: 422–427.
- Assum, G., T. Fink, C. Klett, B. Lengel, M. Schanbacher, S. Uhl, and G. Wöhr. 1991. A new multisequence family in human. *Genomics* 11: 34–41.
- Baban, S., J. Freeman, and D. Mager. 1996. Transcripts from a novel human KRAB zinc finger gene contain spliced Alu and endogenous retroviral segments. *Genomics* 33: 463–472.
- Bailey, T. and C. Elkan. 1994. *Fitting a mixture model by expectation maximization to discover motifs in biopolymers Conference on Intelligent Systems for Molecular Biology*, pp. 28–36. AAAI Press, Menlo Park, CA.
- Batzer, M., S. Arcot, J. Phinney, M. Alegria-Hartman, D. Kass, S. Milligan, C. Kimpton, P. Gill, M. Hochmeister, A. Panayiotis et al. 1996. Genetic variation of recent Alu insertions in the human populations. *J. Mol. Evol.* 42: 22–29.
- Becker, K., J. Nagle, R. Canning, W. Biddison, K. Ozato, and P. Drew. 1995. Rapid isolation and characterization of 118 novel C2H2-type zinc finger cDNAs expressed in human brain. *Hum. Mol. Genet.* 4: 685–691.
- Bellefroid, E., D. Poncelet, P. Lecocq, O. Relevant, and J. Martial. 1991. The evolutionarily conserved Kruppel-associated box domain defines a subfamily of

- eukaryotic multifingered proteins. *Proc. Natl. Acad. Sci.* 88: 3608–3612.
- Bellefroid, E., J.-C. Marine, T. Ried, P. Lecocq, M. Riviere, C. Amemiya, D. Poncet, P. Coulie, P. deJong, C. Szpirer, et al. 1993. Clustered organization of homologous KRAB zinc-finger genes with enhanced expression in human T lymphoid cells. *EMBO J.* 12: 1363–1374.
- Bellefroid, E., J. Marine, A. Matera, C. Bourguignon, T. Desai, K. Healy, P. Bray-Ward, J. Martial, J. Ihle, and D. Ward. 1995. Emergence of the ZNF91 Kruppel-associated box-containing zinc finger gene family in the last common ancestor of anthropoidea. *Proc. Natl. Acad. Sci.* 92: 10757–10761.
- Bishop, J., G. Selman, J. Hickman, L. Black, R. Saunders, and A. Clark. 1985. The 45-kb unit of major urinary protein gene organization is a gigantic imperfect palindrome. *Mol. Cell. Biol.* 5: 1591–1600.
- Brandriff, B., L. Gordon, and B. Trask. 1991. A new system for high-resolution DNA sequence mapping interphase pronuclei. *Genomics* 10: 75–82.
- Brandriff, B., L. Gordon, A. Fertitta, A. Olsen, M. Christensen, L. Ashworth, D. Nelson, A. Carrano, and H. Mohrenweiser. 1994. Human chromosome 19p: A fluorescence in situ hybridization map with genomic distance estimates for 79 intervals spanning 20 Mb. *Genomics* 23: 582–591.
- Carrano, A., P. de Jong, E. Branscomb, T. Slezak, and B. Watkins. 1989. Constructing chromosome- and region-specific cosmid maps of the human genome. *Genome* 31: 1059–1065.
- Constantinou-Deltas, C., E. Bashiardes, P. Patsalis, M. Hadjimarou, P. Kroisel, P. Ioannou, A. Roses, and J. Lee. 1996. Complete coding sequence, exon/intron arrangement and chromosome location of ZNF45, a KRAB-domain-containing gene. *Cytogenet. Cell Genet.* 75: 230–233.
- Cooper, K., R. Fisher, and C. Tyler-Smith. 1992. Structure of the pericentric long arm region of the human Y chromosome. *J. Mol. Biol.* 228: 421–432.
- de Jong, P., K. Yokabata, C. Chen, F. Lohman, L. Pederson, J. McNinch, and M. Van Dilla. 1989. Human chromosome-specific partial digest libraries in lambda and cosmid vectors. *Cytogenet. Cell Genet.* 51: 985.
- Derry, J., U. Jess, and U. Francke. 1995. Cloning and characterization of a novel zinc finger gene in Xp11.2. *Genomics* 30: 361–365.
- Devine, S. and J. Boeke. 1994. Efficient integration of artificial transposons into plasmid targets *in vitro*: A useful tool for DNA mapping, sequencing and genetic analysis. *Nucleic Acids Res.* 22: 3765–3772.
- Di Cristofano, A., M. Strazzullo, L. Longo, and G. La Mantia. 1995. Characterization and genomic mapping of the ZNF80 locus: Expression of this zinc-finger gene is driven by a solitary LTR of ERV9 endogenous retroviral family. *Nucleic Acids Res.* 23: 2823–2830.
- Eichler, E., F. Lu, Y. Shen, R. Antonacci, V. Jurecic, N. Doggett, R. Moyzis, A. Baldini, R. Gibbs, and D. Nelson. 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.
- Eichler, E., M. Budarf, M. Rocchi, L. Deaven, N. Doggett, A. Baldini, D. Nelson, and H. Mohrenweiser. 1997. Interchromosomal duplications of the adrenoleukodystrophy locus: A phenomenon of pericentromeric plasticity. *Hum. Mol. Genet.* 6: 991–1002.
- Friedman, J., W. Fredericks, D. Jensen, D. Speicher, X. Huang, E. Neilson, and F. Rauscher. 1996. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes & Dev.* 10: 2067–2078.
- Gao, L., M. Frey, and A. Matera. 1997. Human genes encoding U3 snRNA associate with coiled bodies in interphase cells and are clustered on chromosome 17p11.2 in a complex inverted repeat structure. *Nucleic Acids Res.* 25: 4740–4747.
- Greig, G. and H. Willard. 1992. Beta satellite DNA: Characterization and localization of two subfamilies from the distal and proximal short arms of human acrocentric chromosomes. *Genomics* 12: 573–580.
- Grondin, B., M. Bazinet, and M. Aubry. 1996. The KRAB zinc finger gene ZNF74 encodes an RNA-binding protein tightly associated with the nuclear matrix. *J. Biol. Chem.* 271: 15458–15467.
- Groot, P., W. Mager, N. Henriquez, J. Pronk, F. Arwert, R. Planta, A. Eriksson, and R. Frants. 1990. Evolution of the human alpha-amylase multigene family through unequal, homologous and inter- and intrachromosomal crossover. *Genomics* 8: 97–105.
- Hoffman, S., R. Hromas, C. Amemiya, and H. Mohrenweiser. 1996. The location of MZF-1 at the telomere of human chromosome 19q makes it vulnerable to degeneration in aging cells. *Leuk. Res.* 20: 281–283.
- Hoovers, J., M. Mannens, R. John, J. Blik, V. van Heynigen, D. Poreous, N. Leschot, A. Westerveld, and P. Little. 1992. High-resolution localization of 69 potential human zinc finger protein genes: A number are clustered. *Genomics* 12: 254–263.
- Hromas, R., S. Collins, D. Hickstein, W. Raskind, L. Deaven, P. O'Hara, F. Hagen, and K. Kaushansky. 1991. A retinoic acid-responsive human zinc finger gene, MZF-1, preferentially expressed in myeloid cells. *J. Biol. Chem.* 266: 14183–14187.
- Huebner, K., T. Druck, C. Croce, and H. Thiesen. 1991. Twenty-seven nonoverlapping zinc finger cDNAs from human T cells map to nine different chromosomes with apparent clustering. *Am. J. Hum. Genet.* 48: 726–740.

- Hyrien, O., M. Debatisse, G. Buttin, and B. de Saint Vincent. 1988. The multicopy appearance of a large inverted duplication and the sequence at the inversion joint suggest a new model for gene amplification. *EMBO J.* 7: 407–417.
- Jackson, M., C. See, L. Mulligan, and B. Lauffart. 1996. A 9.75-Mb map across the centromere of human chromosome 10. *Genomics* 33: 258–270.
- Kim, S., Y. Chen, E. O'Leary, R. Witzgall, M. Vidal, and J. Bonventre. 1996. A novel member of the RING finger family, KRIP-a, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc. Natl. Acad. Sci.* 93: 15299–15304.
- Klug, A. and J. Schwabe. 1995. Zinc fingers. *FASEB J.* 9: 597–604.
- Kristensen, T., H. Voss, and W. Ansorge. 1987. A simple and rapid preparation of M13 templates for manual and automated dideoxy sequencing. *Nucleic Acids Res.* 15: 5507–5516.
- Lamerdin, J., S. Stilwagen, M. Ramirez, L. Stubbs, and A. Carrano. 1996. Sequence analysis of the ERCC2 gene region in human, mouse and hamster reveals three linked genes. *Genomics* 34: 399–409.
- Lichter, P., P. Bray, T. Ried, I. Dawid, and D. Ward. 1992. Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile sites of human chromosomes. *Genomics* 13: 999–1007.
- Margolin, J., J. Friedman, W. Meyere, H. Vissing, H. Thiesen, and F. Rauscher. 1994. Kruppel-associated boxes are potent transcriptional repression domains. *Proc. Natl. Acad. Sci.* 91: 4509–4513.
- Miller, J., A. McLachan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIA from *Xenopus oocytes*. *EMBO J.* 4: 1609–1614.
- Mohrenweiser, H., A. Olsen, A. Archibald, C. Beattie, M. Burmeister, J. Lamerdin, G. Lennon, E. Stewart, L. Stubbs, J. Weber et al. 1996. Report on abstracts of the third international workshop on human chromosome 19 mapping 1996. *Cytogenet. Cell Genet.* 74: 161–186.
- Munzy, D., S. Richards, Y. Shen, and R. Gibbs. 1993. PCR based strategies for gap closure in large scale sequencing projects. In *Automated DNA sequencing and analysis techniques* (ed. C. Venter). Harcourt, Brace, and Jovanovich, London, UK.
- Ogawa, T., D. Poncet, Y. Kinoshita, T. Noce, M. Takeda, K. Kawamoto, K. Udagawa, P. Lecocqu, J. Marine, J. Martial et al. 1998. Enhanced expression in seminoma of human zinc finger genes located on chromosome 19. *Canc. Genet. Cytogenet.* 100: 36–42.
- Parrish, J., E. Eichler, T. Shofield, A. Chinault, M. Graves, A. Arenson, C. Lee, and D. Nelson. 1995. Cosmid binning and cDNA identification in Xq28. *Am. J. Hum. Genet.* 57: A267.
- Parsons, J. 1995. Miropeats: Graphical DNA sequence comparisons. *Comput. Appl. Biosci.* 11: 615–619.
- Parsons, R., G.-M. Li, M.J. Longley, W.-H. Fang, N. Papadopoulos, J. Jen, A. de la Chapelle, K.W. Kinzler, B. Vogelstein, and P. Modrich. 1993. Hypermutability and mismatch repair deficiency in RER+ tumor cells. *Cell* 75: 1227–1236.
- Pengue, G., A. Caputo, C. Rossi, G. Barbanti-Brodano, and L. Lania. 1995. Transcriptional silencing of human immunodeficiency virus type 1 long terminal repeat-driven gene expression by the Kruppel-associated box repressor domain targeted to the transactivating response element. *J. Virol.* 69: 6577–6580.
- Pieler, T. and E. Bellefroid. 1994. Perspectives on zinc finger protein function and evolution—an update. *Mol. Biol. Rep.* 20: 1–8.
- Regnier, V., M. Meddeb, G. Lecointre, F. Richard, A. Duverger, V. Nguyen, B. Dutrillaux, A. Bernheim, and G. Danglot. 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.
- Rosenberg, U.B., C. Schroeder, A. Kienlin, S. Cote, I. Riede, and H. Jaekle. 1986. Molecular genetics of Kruppel, a gene required for segmentation of the *Drosophila* embryo. *Nature* 319: 336–339.
- Rousseau-Merck, M., J. Hillion, P. Jonveaux, P. Couillin, P. Seite, H. Thiesen, and R. Berger. 1993. Chromosomal localization of 9 KOX zinc finger genes: Physical linkages suggest clustering of KOX genes on chromosomes 12, 16 and 19. *Hum. Genet.* 92: 583–587.
- Shannon, M., L. Ashworth, M. Mucenski, J. Lamerdin, E. Branscomb, and L. Stubbs. 1996. Comparative analysis of a conserved zinc finger gene cluster on human chromosome 19q and mouse chromosome 7. *Genomics* 33: 112–120.
- Shen, M.-R., M. Batzer, and P. Deininger. 1991. Evolution of the master Alu gene(s). *J. Mol. Evol.* 33: 311–320.
- Shiels, C., C. Coutelle, and C. Huxley. 1997. Contiguous arrays of satellites 1, 3, and beta form a 1.5-Mb domain on chromosome 22p. *Genomics* 44: 35–44.
- Smit, A. 1993. Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21: 1863–1872.
- Smit, A. and A. Riggs. 1995. MIRs are classic, tRNA-derived SINES that amplified before the mammalian radiation. *Nucleic Acids Res.* 23: 98–102.
- Smit, A., G. Toth, A. Riggs, and J. Jurka. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol.* 246: 401–417.
- Sonnhammer, E. and R. Durbin. 1995. A dot matrix program with dynamic threshold control suited for

- genomic DNA and protein sequence analysis. *Gene* 167: GC1-10.
- Teglund, S., A. Olsen, W. Khan, L. Frangsmyr, and S. Hammarstrom. 1994. The pregnancy-specific glycoprotein (PSG) gene cluster on human chromosome 19: Fine structure of the 11 PSG genes and identification of 6 new genes forming a third subgroup within the carcinoembryonic antigen (CEA) family. *Genomics* 23: 669-684.
- Thiesen, H., E. Bellefroid, O. Relevant, and J. Martial. 1991. Conserved KRAB protein domain identified upstream from the zinc finger region of Kox 8. *Nucleic Acids Res.* 19: 3996.
- Tommerup, N. and H. Vissing. 1995. Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders. *Genomics* 27: 259-264.
- Tommerup, N., L. Aagaard, C. Lund, E. Boel, S. Baxendale, G. Bates, H. Lehrach, and H. Vissing. 1993. A zinc-finger gene ZNF141 mapping at 4p16.3/D4S90 is a candidate gene for Wolf-Hirschhorn (4p-) syndrome. *Hum. Mol. Genet.* 2: 1571-75.
- Trask, B., A. Fertitta, M. Christensen, J. Youngblom, A. Bergmann, A. Copeland, P. de Jong, H. Mohrenweiser, A. Olsen, A. Carrano et al. 1993. Fluorescence in situ hybridization mapping of human chromosome 19: Cytogenetic band location of 540 cosmids and 70 genes or DNA markers. *Genomics* 15: 133-145.
- Trask, B., C. Friedman, A. Martin-Gallardo, L. Rowen, C. Akinbami, J. Blankenship, C. Collins, D. Giorgi, S. Iadonato, F. Johnson et al. 1998. Members of the olfactory receptor gene family are contained in large blocks of DNA duplicated polymorphically near the ends of human chromosomes. *Hum. Mol. Genet.* 1998: 13-26.
- Tunnacliffe, A., L. Liu, J. Moore, M. Leversha, M. Jackson, L. Papi, M. Ferguson-Smith, H. Thiesen, and B. Ponder. 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.
- Uberbacher, E. and R. Mural. 1991. Locating protein-coding regions in human DNA sequences by multiple sensor-neural network approach. *Proc. Natl. Acad. Sci.* 88: 11261-11265.
- Villa, A., I. Zucchi, G. Pilia, D. Strina, L. Susani, F. Morali, C. Patrosso, A. Frattini, F. Lucchini, and M. Repetto. 1993. ZNF75: Isolation of a cDNA clone of the KRAB zinc finger gene subfamily mapped in YACs 1 Mb telomeric of HPRT. *Genomics* 18: 223-229.
- Vissing, H., W. Meyere, L. Aagaard, N. Tommerup, and H. Thiesen. 1995. Repression of transcriptional activity by heterologous KRAB domains present in zinc finger proteins. *FEBS Lett.* 369: 153-157.
- 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.
- Waye, J. and H. Willard. 1989. Human beta satellite DNA: Genomic organization and sequence definition of a class of highly repetitive tandem DNA. *Proc. Natl. Acad. Sci.* 86: 6250-6254.
- Wilkinson, D., S. Bhatt, P. Chavrier, R. Bravo, and P. Charnay. 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* 337: 461-464.
- Willard, H. 1990. Alpha and beta satellite sequences on chromosome 21: The possible role of centromere and chromosome structure in nondisjunction. *Prog. Clin. Biol. Res.* 360: 39-52.
- Windle, B. and G. Wahl. 1992. Molecular dissection of mammalian gene amplification: New mechanistic insights revealed by analyses of very early events. *Mutation Res.* 276: 199-224.
- Winokur, S., U. Bengtsson, J. Vargas, J. Wasmuth, M. Altherr, B. Weiffenbach, and S. Jacobsen. 1996. The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* 5: 1567-1575.
- Witzgall, R., E. O'Leary, A. Leaf, D. Onaldi, and J. Bonventre. 1994. The Kruppel-associated box-A (KRAB-A) domain of zinc finger proteins mediates transcriptional repression. *Proc. Natl. Acad. Sci.* 91: 4514-4518.
- Wohr, G., T. Fink, and G. Assum. 1996. A palindromic structure in the pericentromeric region of various human chromosomes. *Genome Res.* 6: 267-279.
- Worton, R., J. Sutherland, J. Sylvester, H. Willard, S. Bodrug, I. Dube, C. Duff, V. Kean, P. Ray, and R. Schmickel. 1988. Human ribosomal RNA genes: Orientation of the tandem array and conservation of the 5' end. *Science* 239: 64-68.
- Zimonjic, D., M. Kelley, J. Rubin, S. Aaronson, and N. Popescu. 1997. Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and humans. *Proc. Natl. Acad. Sci.* 94: 11461-11465.

Received April 17, 1998; accepted in revised form June 23, 1998.