The EMBO Journal vol.14 no.1 pp.169–177, 1995

Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the *RNU2* locus

Thomas Pavelitz, Laura Rusché, A.Gregory Matera[1,2], Jeremiah M.Scharf and Alan M.Weiner[3]

Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06510 and [1]Department of Genetics, Yale University School of Medicine, New Haven, CT 06510, USA
[2]Present address: Department of Genetics, Case Western Reserve School of Medicine, Cleveland, OH 44106-4955, USA

[3]Corresponding author

Communicated by A.J.Jeffreys

In primates, the tandemly repeated genes encoding U2 small nuclear RNA evolve concertedly, i.e. the sequence of the U2 repeat unit is essentially homogeneous within each species but differs somewhat between species. Using chromosome painting and the *NGFR* gene as an outside marker, we show that the U2 tandem array (*RNU2*) has remained at the same chromosomal locus (equivalent to human 17q21) through multiple speciation events over >35 million years leading to the Old World monkey and hominoid lineages. The data suggest that the U2 tandem repeat, once established in the primate lineage, contained sequence elements favoring perpetuation and concerted evolution of the array *in situ*, despite a pericentric inversion in chimpanzee, a reciprocal translocation in gorilla and a paracentric inversion in orang utan. Comparison of the 11 kb U2 repeat unit found in baboon and other Old World monkeys with the 6 kb U2 repeat unit in humans and other hominids revealed that an ancestral U2 repeat unit was expanded by insertion of a 5 kb retrovirus bearing 1 kb long terminal repeats (LTRs). Subsequent excision of the provirus by homologous recombination between the LTRs generated a 6 kb U2 repeat unit containing a solo LTR. Remarkably, both junctions between the human U2 tandem array and flanking chromosomal DNA at 17q21 fall within the solo LTR sequence, suggesting a role for the LTR in the origin or maintenance of the primate U2 array.

Key words: concerted evolution/gene amplification/solo LTR

Introduction

Tandemly repeated multigene families are common among eukaryotes, some of the best studied families being those encoding the large ribosomal RNAs, 5S RNA, small nuclear RNAs, tRNAs and histones. One virtue of a multigene family is obvious: multiple gene copies allow higher levels of gene expression when transcription of a single gene copy cannot meet the needs of the cell. What is far from obvious, however, is why tandemly repeated multigene families are characteristically homogeneous, and what mechanisms might be responsible for maintaining this homogeneity (Edelman and Gally, 1970; Weiner and Denison, 1983; Dover, 1993; Jinks-Robertson and Petes, 1993). The concerted evolution of tandemly repeated genes cannot reflect selection at the level of the gene product because intergenic regions are no less homogeneous than the coding regions, and the coding regions do not diverge, although many mutations would be silent or selectively neutral. Concerted evolution of tandemly repeated multigene families must therefore reflect mechanisms that maintain the homogeneity of the tandem repeat as DNA but are substantially blind to the gene sequences within it. These sequence turnover mechanisms can work with remarkable efficiency at the population level, rapidly replacing one tandemly repeated sequence with a closely related variant sequence (Dover, 1993; Elder and Turner, 1994). The evolution of alphoid (Warburton *et al.*, 1993) and other non-coding satellite sequences (Hipeau-Jacquotte *et al.*, 1989), as well as mammalian minisatellites (Armour and Jeffreys, 1992; Armour *et al.*, 1993; Jeffreys *et al.*, 1994), suggests that concerted evolution may be a necessary consequence of tandemly repeated sequence organization, but it is also possible that specific sequences within (or flanking) the repeat unit may favor, disfavor or even be required for concerted evolution. Concerted evolution of a tandemly repeated gene family also has profound genetic consequences: to the extent that all gene copies within an array are constrained to be identical, the multigene family is genetically transformed into a single copy gene.

The tandemly repeated human U2 genes appear to be an excellent system for studying the characteristics and, ultimately, the mechanism of concerted evolution in mammals. We have shown previously that the human U2 tandem array is essentially homogeneous (Van Arsdell and Weiner, 1984a; Westin *et al.*, 1984) and maps to a single chromosomal locus (*RNU2*) at 17q21 (Lindgren *et al.*, 1985a; Hammarstrom *et al.*, 1985). This contrasts, for example, with mammalian ribosomal RNA genes which have a much larger repeat unit (>43 versus 6 kb), are grouped in multiple non-syntenic clusters (nucleolus organizers), and are highly polymorphic both within and between chromosomes (Seperack *et al.*, 1988; Gonzalez *et al.*, 1992). Although the tandem repeat unit of the human 5S ribosomal RNA genes (2.2 kb) is even smaller than the U2 repeat unit, the 5S genes are highly polymorphic, map to multiple sites (Sorensen *et al.*, 1991), and are difficult to distinguish from an abundance of pseudogenes (Sorensen and Frederiksen, 1991). An additional appeal of the human U2 genes is that the *RNU2* locus colocalizes with an adenovirus 12-inducible fragile site (Lindgren *et al.*, 1985a; Durnam *et al.*, 1988) and ectopic U2 tandem arrays, introduced by gene transfer

techniques, generate new virally inducible fragile sites (Li *et al.*, 1993). Taken together, these data suggest that sequences within the U2 array have the potential to affect chromosome structure, stability and recombination.

We have also shown that the U2 tandem array, once established early in the primate lineage, has been stable over >35 million years and has evolved in a concerted fashion (Matera *et al.*, 1990). Specifically, Old World monkeys (baboon, macaque and talapoin) have an 11 kb U2 repeat unit, while apes (gibbon, orang utan, chimpanzee, gorilla) and humans have a 6 kb U2 repeat unit, but it was not clear whether the 5 kb difference was due to deletion in one lineage or insertion in the other. Here we show that concerted evolution of the primate U2 tandem repeat occurs without cytologically detectable movement of the array, a result which tends to rule out models for concerted evolution involving excision and re-integration of repeat units (Bernstein *et al.*, 1985; Lindgren *et al.*, 1985b; Matera *et al.*, 1990). We also show that the 5 kb difference between Old World monkey and hominoid U2 repeat units is due to homologous excision of a 6 kb provirus, leaving behind a solo LTR, and that both junctions between the U2 tandem array and flanking chromosomal DNA fall within the solo LTR. These data demonstrate that concerted evolution can spread deletions of >5 kb from one repeat unit to all other copies, and may implicate the proviral LTR in the origin or maintenance of the U2 tandem array.

Results

The cytological context of U2 genes has been conserved through multiple speciation events

Concerted evolution of the primate *RNU2* locus (Matera *et al.*, 1990) is consistent with two very different scenarios: either the U2 genes are evolving *in situ* (perhaps by unequal sister chromatid exchange and/or gene conversion) or the U2 genes are repeatedly excised and re-integrated at new loci (with the excised copy serving as the founder for the next cycle of amplification). To distinguish between these two scenarios, we have used fluorescence *in situ* hybridization on human, chimpanzee, orang utan and baboon metaphase chromosome spreads. Homologous chromosomes were identified by hybridization with probe sets prepared from a human chromosome 17 library ('chromosome painting'). The U2 tandem array (*RNU2* locus) was detected using the intact U2 repeat as probe; truncated U2 retropseudogenes scattered throughout the genome are too short to react significantly with this probe (Hammarstrom *et al.*, 1984; Van Arsdell and Weiner, 1984b; Lindgren *et al.*, 1985a) and the presence of competitor C_0t 1 DNA suppresses signals from more highly repetitive probe sequences (Wienberg *et al.*, 1990). A cosmid probe for the human nerve growth factor receptor gene (*NGFR*), located 8–10 map units more proximal to the telomere than *RNU2* in humans (A.J.Pakstis and K.Kidd, personal communication), was used to establish that the immediate environment of *RNU2* was unchanged by speciation.

As shown in Figure 2A, the *RNU2* locus resides on the chromosome 17 homolog in each of the three species examined, i.e. on baboon 17q, on orang utan 19q and on chimpanzee 19p as expected for a pericentric inversion

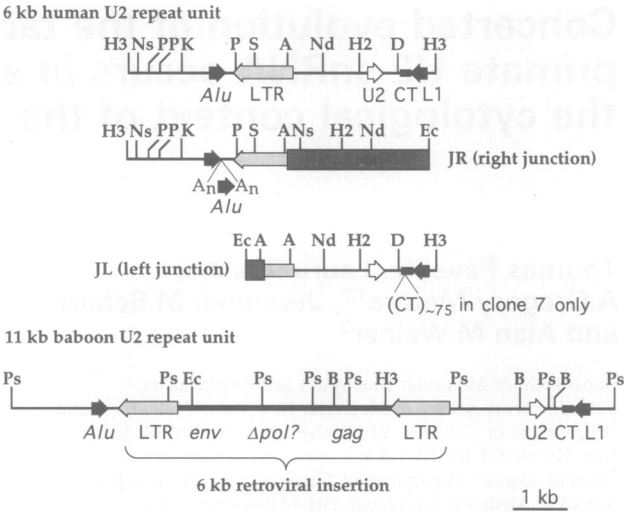

Fig. 1. Maps of the human U2 repeat unit, the junction fragments with flanking DNA and the baboon U2 repeat unit. The maps are drawn approximately to scale, and the human junction fragments JL and JR are aligned with the human U2 repeat to emphasize the location of the junctions within the solo LTR. The U2 RNA coding region, Alu elements, a 5′ and 3′ truncated L1 element and a nearly perfect dinucleotide repeat of $(CT)_{70}$ are indicated. As shown, both JR fragments contain an additional Alu element and one JL contains an additional $(CT)_{\sim 75}$, compared with the U2 repeat unit (see Figure 3B and text). The indicated features of the baboon U2 repeat unit have all been verified by DNA sequence analysis; *gag* appears to be immediately upstream from *env*, suggesting that *pol* may be deleted. A, *Ase*I; B, *Bam*HI; D, *Dra*I; Ec, *Eco*RI; H2, *Hin*cII; H3, *Hin*dIII; K, *Kpn*I; Nd, *Nde*I; Ns, *Nsi*I; Ps, *Pst*I; P, *Pvu*II; S, *Spe*I. JL, JR and the baboon U2 repeat have not yet been mapped with *Hin*cII or *Pvu*II.

(Yunis and Prakash, 1982; Wienberg *et al.*, 1990). *RNU2* was already known to reside on gorilla 4p (Matera and Marks, 1993), the homolog of human 17q (Wienberg *et al.*, 1990). As shown in Figure 2B, the local chromosomal context of *RNU2* has been conserved over this evolutionary period because the *RNU2* and *NGFR* loci are as tightly associated in baboon as in human although the map order of the two loci is reversed in baboon, consistent with the previously hypothesized paracentric inversion (Yunis and Prakash, 1982; Matera and Marks, 1993). *RNU2* is also more terminal in both baboon and orang utan, as expected for a paracentric inversion.

Cloning of junction fragments

To facilitate these studies, the sequence of a complete U2 repeat unit was determined (GenBank accession No. L37793) and the 5834 bp sequence was arbitrarily numbered from the unique *Hin*dIII site, as shown schematically in Figure 1. Gene counting experiments had originally suggested that each *RNU2* locus would contain ~10 tandem copies of the 6 kb repeat unit (Van Arsdell and Weiner, 1984; Westin *et al.*, 1984). Digestion of genomic DNA with restriction enzymes such as *Eco*RI that do not cut within the U2 repeat unit ('null cutters') should therefore generate an unusually large restriction fragment corresponding to the intact array with attached junctions. These junction fragments are provisionally designated JL and JR (for left and right) until the orientation of the U2 array on 17q can be determined. Intact U2 arrays were prepared by field inversion gel electrophoresis (FIGE) of *Eco*RI digested HT1080 genomic DNA under conditions that

intentionally did not resolve the two parental arrays. Purification of the arrays by this technique should approach 10^5-fold, but incomplete digestion apparently limits actual enrichment to ~10^3-fold. The enriched *Eco*RI U2 arrays were redigested with a variety of enzymes, blotted and probed with the 3.7 kb *Pvu*II−*Hin*dIII fragment from the right end of the U2 repeat unit (Figure 1). These blots revealed a single fragment whose size in different restriction digests varied in a predictable fashion consistent with a position at the left end of the array (data not shown). This suggested that JL could be cloned as an *Eco*RI−*Hin*dIII fragment of ~3.5 kb. We were unable to detect a comparable *Hin*dIII−*Eco*RI fragment spanning JR, but we did identify a unique 3.6 kb fragment when the intact *Eco*RI U2 array was digested with *Nsi*I and probed with the 584 bp *Hin*dIII−*Pvu*II fragment from the left end of the U2 repeat unit (data not shown). Nonetheless, we attempted to clone both JL and JR as *Hin*dIII−*Eco*RI fragments derived from the enriched *Eco*RI array. Of three independent candidate clones (clones 11, 7 and 17), all proved to be bona fide junctions (see below). In particular, it became clear that we had failed to detect JR in blotting experiments because the *Hin*dIII−*Eco*RI and *Nsi*I−*Eco*RI fragments spanning JR comigrate with the multicopy U2 repeat unit excised by *Hin*dIII or *Nsi*I (Figure 1); the 3.6 kb *Nsi*I band we detected was an internal fragment of clone 11 (Figure 1).

Verification of junction fragments

The maps of clones 7, 11 and 17 suggested that all three were junction fragments; the *Hin*dIII end of each clone resembles the U2 repeat but the *Eco*RI end is completely divergent, as expected if the *Hin*dIII ends derive from the tandem array and the *Eco*RI ends from flanking chromosomal DNA. Thus clone 11 is likely to be JR, clones 7 and 17 alleles of JL. As discussed below, the 0.3 kb present in clone 11 but not in the corresponding region of the U2 repeat is due to insertion of an additional Alu element into the poly(A) tract of the U2 Alu (Figure 1); the ~150 bp present in clone 7 but not in clone 17 or in the U2 repeat reflects expansion of the $(CT)_{70}$ array to $(CT)_{>140}$. The identity of these three clones as bona fide junction fragments was confirmed in two ways. First, clone 11 was shown by *in situ* hybridization to colocalize cytologically with the U2 repeat (Figure 2C and D). Second, both JR and JL were shown by field inversion gel electrophoresis to comigrate with the intact U2 arrays (Figure 3). To do this, we took advantage of the observation that the parental U2 arrays of HT1080 cells could be resolved after digestion of genomic DNA with the null cutter *Eco*RI (Figure 3A); redigestion of the separated arrays with *Hin*dIII indicated that clone 17 corresponds to JL of the upper array, clone 7 to JL of the lower array and clone 11 to JR of both arrays (Figure 3B; data not shown for JR). The two HT1080 arrays are ~75 and 89 kb when excised with *Eco*RI; the *Eco*RI−*Hin*dIII fragments spanning JL and JR together account for 9.5 kb, so the two arrays contain ~11 and 13 copies of the U2 repeat unit respectively.

Sequence analysis of junction fragments

The U2 tandem array breaks off abruptly at the left junction with 17q21 at position 2597 of the U2 repeat

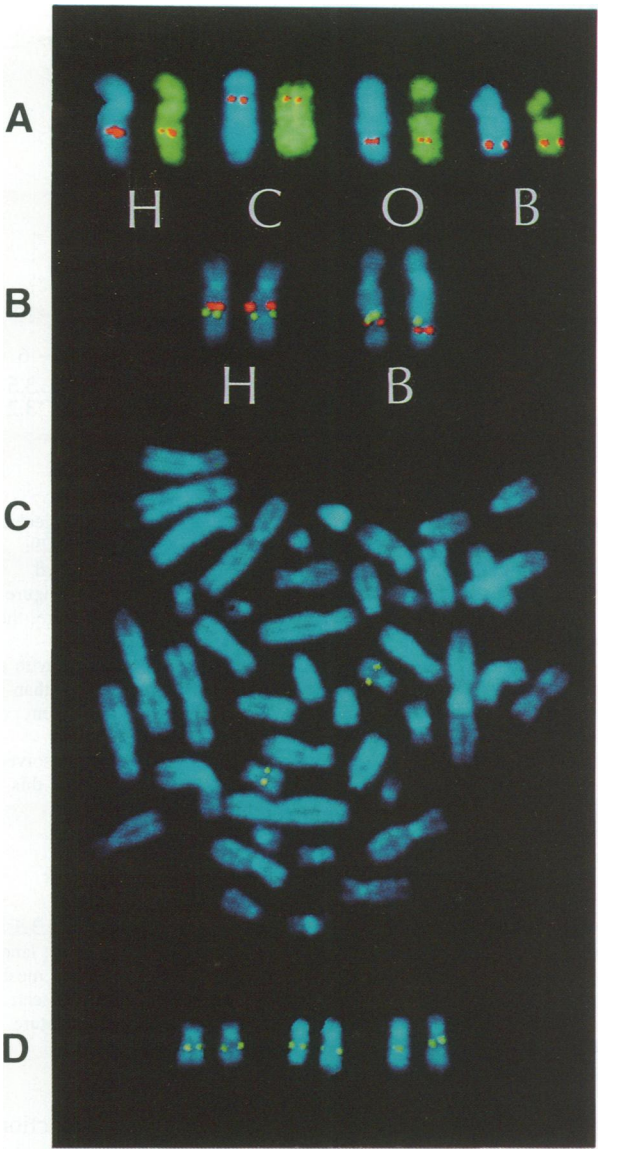

Fig. 2. Cytogenetic localization of U2 genes in Old World monkeys and hominids by fluorescence *in situ* hybridization. (**A**) *RNU2* maps to the chromosome 17 homolog in baboon, orang utan, gorilla and chimpanzee: yellow, human chromosome 17 paint; blue, DAPI stain for DNA; red, *RNU2* locus. (**B**) *NGFR* and *RNU2* are as tightly associated in baboon as in human, although the map order is reversed: red and blue as in (A); yellow, *NGFR* locus. (**C**) Representative metaphase spread probed with JR. (**D**) Three pairs of chromosomes 17 excerpted from metaphase spreads as shown in (C): blue, DAPI; yellow, JR. H, human; C, chimpanzee; O, orang utan; B, baboon.

sequence (Figure 4A). In contrast, the array breaks off stepwise at the right junction between positions 2738 and 3047 of the U2 repeat sequence; homology between JR and the U2 repeat sequence decays from excellent (2553–2747) to moderate (2748–3047) to undetectable (>3048) with an apparent insertion of 52 bp in JR at U2 position 2738 and a deletion of 24 bp in JR at U2 position 2851 (Figure 4B). The 52 bp insertion is in fact a nearly perfect tandem duplication of the previous 52 bp (D.Liao, unpublished data). The stepwise decay of homology at JR, accompanied by internal duplication and deletion, suggests that this junction has been remodeled more than once: it is difficult to imagine that any single recombination

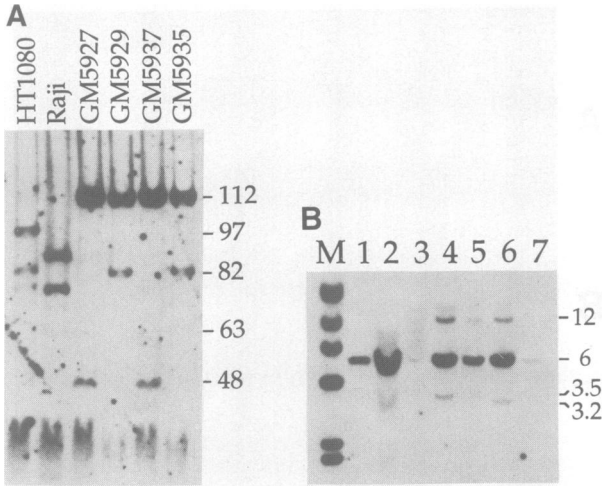

Fig. 3. JR and JL comigrate with intact U2 arrays. (**A**) *Eco*RI digested DNA from six human cell lines (HT1080, Raji, GM5927, GM5929, GM5937 and GM5935) was resolved by FIGE, blotted and probed with the 3.7 kb *Pvu*II–*Hin*dIII fragment of the U2 repeat unit (Figure 1). Intact parental U2 arrays are visible as bands of 48 kb or more; the smear below represents background hybridization to bulk *Eco*RI fragments. The HT1080 cells used here contain a subpopulation with a novel U2 tandem array that is apparently one repeat unit shorter than the smaller of the two parental arrays; most subclones derived from this mixed HT1080 population lack the shorter array (A.D.Bailey, unpublished data). (**B**) An *Eco*RI digest of HT1080 DNA was resolved by preparative FIGE as in (A) and successive size fractions from this gel were redigested with *Hin*dIII, resolved by agarose gel electrophoresis, blotted and probed with the unique *Eco*RI–*Ase*I fragment from the left end of JL (lanes 4–8) or the *Nsi*I–*Eco*RI fragment from the right end of JR (data not shown). As shown in Figure 1, the JL probe reacts with both left junctions; these two *Eco*RI–*Hin*dIII fragments, corresponding to clones 7 and 17, are 3.5 and 3.2 kb in length and differ by $(CT)_{\sim 75}$. Lane 4, smaller array; lane 5, DNA between the arrays; lane 6, larger array. The 12 kb band must be due to incomplete digestion of the tandem array because it is seen only with DNA which has been eluted from low melting temperature agarose. M, *Hin*dIII digest of λ DNA. Fragment size in kb.

event could generate so many changes in the junction region relative to the intact U2 repeat unit. Surprisingly, both the right and left junctions of the human U2 tandem array with chromosome 17q21 fall within the solo LTR sequence (Figure 1 and see below).

The human U2 repeat contains a solo LTR resulting from homologous excision of a provirus

We assumed initially that the 5 kb difference between the 11 kb U2 repeat unit found in Old World monkeys (baboon, macaque and talapoin) and the 6 kb repeat unit characteristic of apes (gibbon, orang utan, chimpanzee, gorilla) and humans was due to deletion of 5 kb from an ancestral 11 kb repeat in the lineage leading to the apes, or insertion of 5 kb into an ancestral 6 kb repeat in the lineage leading to Old World monkeys. Neither is the case. Instead, as depicted schematically in Figure 1, an ancestral U2 repeat unit was expanded by insertion of a 6 kb retroviral sequence bearing 1 kb LTRs. Subsequent excision of the provirus by homologous recombination between two LTRs generated a founder 6 kb U2 repeat unit containing a 1 kb solo LTR; the 5 kb deletion was then spread to every U2 repeat unit in the array.

The first indication of a proviral insertion within the baboon U2 repeat unit came from DNA sequence analysis;

```
A. left junction

2544 GCCCGGCCTGGCTAATGTCAGACCTCCCACAAGAGGTGGTGGAGCGGAGC 2593
     |    | || |||     ||| | | |  | ||  | ||   || ||  |
 -45 G...AGTCTTGCT....CCAGTCTTGCTCCCAGCCG.GGCACAGTGGCTC -4

2594 GTTCTCTGTCTCCCCTGGAGAGAGGGAGATT.CCTTTCCGGGTCTGCTAA 2642
     |||||||||| ||||||||||||||||||| ||||||||||||||||||
  -3 ACACTCTGTCTCTCCTGGAGAGAGGGAGATTCCCTTTCCGGGTCTGCTAA +47


B. right junction

 2703 GGGCCTTCCCAGGCACTGGCATTACCGCTAGGCCAA.............. 2738
      |||||||||||||||||||||||||||||||| ||||
 -366 GGGCCTTCCCAGGCACTGGCATTACCGCTAGACCAAAGTGTTCTAAATAA -317

 2739 .....................................GGAGCCCTCCA 2749
                                           |||||||||| |
 -316 CAGGTGCCTTCCGAGGAAGAGGCACTACCACAAGACCATGGAGCCCTCAA -267

 2750 GCGGCCCTTCTCTGGGCGTGAATGAGGGCTCACACTCTCGTCTTCTGGTC 2799
      || ||| || || |||| |||  |||||||||| ||||| ||||||||||
 -266 GCAGCCGTTATCCGGGCATGACAGAGGGCTCATACTCTTGTCTTCTGGTT -217

 2800 ACCTCTCACTGTGGCCCTTCAGCTCCTAACTCTGTGTGGCCTGGTTTCCC 2849
      |||||||||  || ||| ||  || || ||||||| ||||||| |||  |
 -216 ACCTCTCACATTGTCCCCTCTACTTCTTACTCTGTATGGCCTGTTTTTTC -167

 2850 CCAAGGTAATCATAATAGAACAGAGATCATTATGGTAATAGAACAAAGAG 2899
      |                          |  | ||  | ||||||||
 -166 C............................TCGGTTATAATAACAAAGAT -146

 2900 TGATGCTACAAACTAATGATTAATAATAGTCAGATATAATCCTATCCGTT 2949
      | || ||| |||||||| ||| ||||||  || || ||||| | || ||
 -145 TAATACTAAAAACTAATAATTGATAATATCCATATGTAATCATCTCTGTA -96

 2950 TCCTATCTCTAGTAAAACTTTTCTTATTCTAATTATTTTCTTTGCTGTAC 2999
      |||||| |||  || ||||||  ||  ||    ||||||||||  | ||
  -95 TCCTATGTCTGATATAACTTTCTTTTATC...CTATTTTCTTTATTATAT -49

 3000 TGGAACAGCTTGTGCCTTCAGGCTCTTGCCTGGGCAGCTCCCTGGCTTGC 3049
      |||||||||||| ||| |||||  |||||||| |||| ||   |||||| |
  -48 TGGAACAGCTTCTGCGTTCAGTGTCTTGCCTTGGCATCTGGATGGCTTTC +2

 3050 GGCCCACAAGATAAGATATATTGCGT.TGAACTATAATTTATGT.TGATT 3097
       |       |     || |  || | ||   | |  | |||||  ||
   +3 TG.TAGGGTGCAGCCCTACAGGGCCTGTGGGTTTTTCTGTATGTGTGCGG +51
```

Fig. 4. DNA sequence of the right and left junctions of the U2 array with flanking chromosomal DNA. Sequence comparison of U2 repeat unit with JL (**A**) and JR (**B**). U2 sequences are numbered from the *Hin*dIII site; junction sequences, derived from the *Ase*I fragment of JL and the *Spe*I–*Nsi*I fragment of JR (Figure 1), are numbered positively to the right of the junction and negatively to the left. Vertical lines denote identity, dots insertions or deletions.

a 1 kb sequence from the 6 kb human U2 repeat unit is present twice within the 11 kb baboon U2 repeat unit, immediately flanking the additional 4 kb sequence found only in the baboon U2 repeat unit (shown schematically in Figure 1). These 1 kb repeats had eluded our preliminary characterization of the baboon repeat by restriction mapping and hybridization with selected probes from the human U2 repeat unit (Matera *et al.*, 1990).

A BLASTN search of GenBank supported the interpretation that the baboon 1 kb repeats are LTRs: the 1 kb sequence from the human U2 repeat (positions 2028–3057) is ~80% homologous to and colinear with nucleotides 2021–3299 of the human cosmid contig HUMHDABCD (McCombie *et al.*, 1992) except for an insertion between nucleotides 2213–2527 of a 300 bp BC200 retrotransposable element (Martignetti and Brosius, 1993) which is absent from both the human and baboon U2 LTRs (Figure 5A). The 1 kb homology in contig HUMHDABCD terminates with 5′ TG...CA 3′ as do all known LTRs, and exhibits no flanking homology with either of the baboon direct repeats, suggesting that it, like the 1 kb sequence in the human U2 repeat unit, represents a solo LTR. Moreover, the uninterrupted colinearity of the human and baboon U2 LTRs (T.Pavelitz, J. A.Leonard and A.M.Weiner, data not shown) indicates that the provirus was excised from the 11 kb Old World monkey repeat unit by homologous recombination between the LTRs.

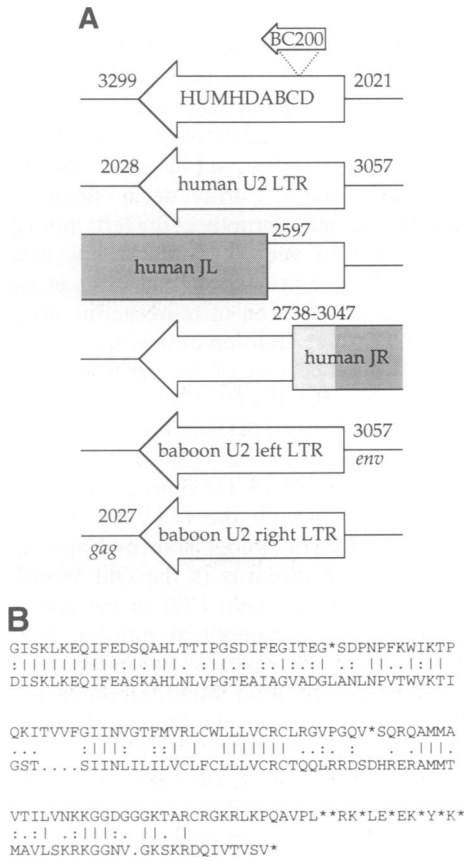

Fig. 5. The human U2 repeat unit contains a solo LTR resulting from homologous excision of the provirus present in the baboon U2 repeat unit. (**A**) Schematic alignment of LTR homologies, drawn approximately to scale. Numbers flanking the HUMHDABCD cosmid contig and human U2 solo LTRs indicate the boundaries of the elements in GenBank coordinates. The BC200 insertion in the HUMHDABCD LTR is indicated. Numbers flanking the baboon LTRs indicate the position in the human U2 repeat sequence where homology breaks off. Numbers above JL and JR show the site (JL) or region (JR) of the junction between the human U2 array (open arrow) and flanking chromosomal DNA (dark shading); the junction region in JR (positions 2738–3047 as shown in Figure 4B) is indicated by light shading. (**B**) Homology between proteins encoded by *env* region of human endogenous retrovirus HERV-K10 (sequence beginning at nt 7922; Ono *et al.*, 1986) and left LTR of baboon U2 provirus. *upper*, baboon LTR; *lower*, HERV-K10. A +1 frameshift was introduced at codon 62 of the baboon sequence to reveal additional homology downstream. Asterisks denote termination codons.

The presumptive proviral insertion in the Old World monkey lineage is at least 35 million years old (Matera *et al.*, 1990), but sequences found between the two LTRs in the baboon U2 repeat unit might still be expected to preserve remnants of retroviral protein coding potential. Indeed, a BLASTN search of GenBank using baboon sequences immediately interior to the presumptive LTRs as the query revealed nucleic acid sequence similarity with the *env* gene of the human endogenous retrovirus HERV-K10 (Ono *et al.*, 1986). This region of nucleic acid similarity was not fortuitous because only one of the six reading frames was nearly free of termination codons; the encoded protein matches the HERV-K10 *env* protein extensively over a region of at least 100 residues (Figure 5B); the open reading frame exhibits numerous third position changes reflecting selection at the protein level (data not shown); and the distance between the amino terminal end of the *env* homology shown in Figure 5B and the downstream LTR is almost exactly the same in the baboon U2 repeat unit (300 nt) as in HERV-K10 (290 nt).

Although the *env* coding region of the baboon U2 provirus is conserved in HERV-K10, homology between the primate U2 and HERV-K10 LTRs is limited to the 3′ most 150 nucleotides (data not shown). The absence of more extensive homology could reflect modular reassortment of LTRs and protein coding sequences (Doolittle and Feng, 1992) or alternatively that LTRs are seldom well conserved because they are assembled from promoter, enhancer and polyadenylation elements which are flexible in sequence and organization.

SINEs and LINEs within the U2 repeat unit

The Alu element within the human U2 repeat bears most of the stigmata of the 'major consensus' element characteristic of Old World monkeys (data not shown) consistent with retroposition of the element into the U2 region of baboon or an immediate ancestor thereof (Schmid and Maraia, 1992; Hellmann-Blumberg *et al.*, 1993). Surprisingly, this Alu lacks flanking direct repeats although it is followed by a pure tract of A_{40}. This may indicate that the poly(A) tract has been maintained or even enlarged despite loss of the presumptive target site duplication, perhaps because the mechanism of concerted evolution favors homogenization of homopolymer and alternating copolymer tracts. An additional Alu element, inserted into the poly(A) tail of the Alu in JR (Figure 1), does not appear to belong to one of the more recently active Alu subfamilies and may reflect retroposition of a less active, highly mutated Alu element (data not shown).

Compared with the active 6 kb L1.2 element (Dombroski *et al.*, 1991), the 483 bp L1 element in the human U2 repeat (positions 5282–5765) is typically 5′ truncated but appears to have undergone secondary mutations after retroposition; it lacks a 3′ terminal poly(A) tract, is not flanked by direct repeats, and is atypically 3′ truncated by 10 bp (data not shown). Given our ability to trace the evolution of the U2 array in deeper primate lineages, a testable hypothesis is that the CT tract is derived from the 3′ terminal poly(T) tract originally generated by retroposition of the L1 element. This situation might resemble the generation of a human minisatellite by internal reduplication of an LTR-like sequence (Armour *et al.*, 1989) or expansion of a short stable repeat in primates into larger hypervariable array in humans (Gray and Jeffreys, 1991).

Discussion

The human U2 genes appear to be an excellent system for studying the origin, stability, concerted evolution and potentially unusual genetic properties of tandemly repeated multigene families in mammals. The tandem array of human U2 genes has a short, essentially homogeneous repeat unit (Van Arsdell and Weiner, 1984a; Westin *et al.*, 1984), maps to a single chromosomal locus *RNU2* (Lindgren *et al.*, 1985a; Hammarstrom *et al.*, 1985), and the concerted evolution of the U2 tandem array can be traced back for >35 million years from hominids to Old World monkeys (Matera *et al.*, 1990). From a more

practical point of view, intact U2 arrays are easily purified by field inversion gel electrophoresis (Figure 3), providing direct experimental access to both parental arrays for studies of *RNU2* recombination, polymorphism and inheritance in humans without recourse to hybrid human/rodent cell lines.

We show here that the U2 tandem array appears to have remained at the same chromosomal locus through multiple speciation events over >35 million years leading to the Old World monkey and hominoid lineages. Admittedly, the resolution of the fluorescence *in situ* hybridization data (Figure 2) is megabases not nucleotides; however, we have recently found that the right junction between the U2 tandem array and flanking chromosomal DNA is nearly identical in chimpanzee, gorilla and human, except for significant remodeling in the junction region itself (Figure 4B and D.Liao and A.M.Weiner, unpublished data). These new data strengthen the case that concerted evolution of the U2 tandem array occurs *in situ* without changing the immediate chromosomal environment of the array. Nonetheless, we cannot exclude the possibility that the entire U2 array, including the flanking sequences, might have undergone local intrachromosomal transposition, one or more times, over the past 35 million years.

Our data tend to rule out models for concerted evolution that require excision of one or more repeat units, followed by re-integration and reamplification of these founder copies elsewhere in the genome. There are at least three reasons for considering such excision/re-integration models. First, U2 genes are dispersed in our prosimian ancestors (Matera *et al.*, 1990). This implies that the primate U2 tandem array was generated by amplification of a single copy prosimian U2 gene and, if amplification can generate the original U2 array, reamplification could maintain it. Second, we have shown previously that an ancient family of human U1 snRNA pseudogenes is clustered at 1q12−q22 on the opposite side of the centromere from the functional U1 snRNA multigene family at 1p36 (Lindgren *et al.*, 1985b). These U1 pseudogenes display extensive flanking homology with the true genes, but the pseudogenes are quite divergent from each other while the true U1 genes are nearly homogeneous. The existence of two related U1 clusters, one old and one new, is prima facie evidence for excision/re-integration of some kind, regardless of whether the old locus gave rise to the new one as we believe (Lindgren *et al.*, 1985b) or both were derived from common ancestral locus. Third, excision/re-integration has been documented experimentally. Amplification of the dihydrofolate reductase gene in CHO cells occurs preferentially on the same chromosome arm, but without loss of the parental single copy locus (Trask and Hamlin, 1989). Such intrachromosomal amplification events could have generated the functional human U1 gene family at 1p36 from the old family at 1q12−q22 (Lindgren *et al.*, 1985b).

The sequence of the right junction region (JR) between the human U2 tandem array and flanking chromosomal DNA suggests that gene conversion plays a role in the concerted evolution of primate U2 genes *in situ*. JR exhibits stepwise loss of homology to the U2 repeat, accompanied by internal duplication and deletion (Figure 4B). While duplication or deletion might reflect reciprocal recombination between homologous or partially homologous sequences, stepwise loss of homology is most consistent with successive rounds of remodeling by gene conversion. Indeed, the structure of JR is reminiscent of the complex gene conversion events seen in human minisatellites (Jeffreys *et al.*, 1994) and the discontinuous gene conversion tracts observed in yeast (Sweetser *et al.*, 1994). Curiously, the U2 array decays stepwise at the right boundary but ends abruptly at the left; this asymmetry might be related to vectorial addition of new coding repeats to the primate involucrin genes (Djian and Green, 1989) or to the polarization of minisatellite arrays where new mutations and the addition of new repeats occur more frequently toward one end of the array (Armour *et al.*, 1993; Jeffreys *et al.*, 1994). The two alphoid satellite junctions characterized to date are abrupt like JL (Jackson *et al.*, 1992; Wevrick *et al.*, 1992).

Comparison of the 11 kb U2 repeat unit in baboon (an Old World monkey) with the 6 kb U2 repeat unit in humans revealed that homologous recombination between the LTRs of a 6 kb provirus in the Old World monkey repeat unit left behind a solo LTR in the ape U2 repeat unit (Figure 1). Thus deletions as large as 5 kb can be spread from one to all copies of the primate U2 tandem array by concerted evolution. More importantly, we found that both junctions between the human U2 tandem array and flanking chromosomal DNA at 17q21 fall within the solo LTR sequence; in fact, the right and left junctions could be as close as 142 bp if the right junction is defined by the proximal insertion at position 2738 (Figure 4B). This suggests a functional role for the LTR in the origin or maintenance of the U2 tandem array, perhaps reflecting the ability of enhancer or promoter elements to confer an open, recombinogenic chromatin structure (Truss *et al.*, 1992; Wu and Lichten, 1993). Indeed, a solo LTR has also been found in the tandem repeat unit of a satellite array in the rodent *Ctenomys* (Pesce *et al.*, 1994). Aging LTRs have been shown to retain residual transcriptional activity in other mammalian systems (Buetti, 1994; Pesce *et al.*, 1994) and the primate U2 LTRs might be expected to do the same; transcription factor binding sites are likely to be more tolerant of mutations than open reading frames, and the open reading frames in the baboon provirus are relatively well preserved (Figure 5B). In the budding yeast *Saccharomyces cerevisiae*, a large number of solo LTRs (called 'solo deltas') are known to affect the regulation of gene expression as well as the course of genomic evolution; these solo deltas are generated by retroposition of Ty elements followed by homologous excision between the LTRs (Fink *et al.*, 1986).

The homogeneity of multigene families is no doubt maintained by multiple mechanisms operating one on top of another (Dover, 1993). Unequal sister and nonsister chromatid exchange could in principle help to maintain the U2 tandem array (Smith, 1974a,b; Armour and Jeffreys, 1992; Warburton *et al.*, 1993) but it has never been shown that the rate of chromatid exchange in any organism is sufficient to eliminate sequence heterogeneity faster than it arises within individual repeat units. Alternatively, the homogeneity of the U2 array might be maintained primarily by the kind of efficient pairwise gene conversion that maintains dispersed multigene families in the yeasts *S.cerevisiae* (Jinks-Robertson and Petes, 1993) and *Schizosaccharomyes pombe* (Amstutz *et al.*, 1985). Gene

conversion between dispersed or imperfectly repeated sequences generally appears to be inefficient in mammals (Bollag *et al.*, 1992; Eikenboom *et al.*, 1994); however, frequencies vary by many orders of magnitude depending on the precise construct (Murti *et al.*, 1992) and tandem DNA sequence organization might stimulate gene conversion, both within and between arrays, by dramatically increasing the local concentration of homologous sequences (Hipeau-Jacquotte *et al.*, 1989). Nor are reciprocal recombination and pairwise gene conversion mutually exclusive mechanisms for maintenance of a tandem array. The ability of reciprocal recombination to cause efficient gene conversion at the crossover point (Orr-Weaver and Szostak, 1985; Jinks-Robertson and Petes, 1993; Curtis *et al.*, 1989) suggests that tandem repetition might also facilitate gene conversion by increasing the frequency of reciprocal recombination. Any or all of these mechanisms could propagate newly arising sequence variants within an array, ultimately generating the kind of complex gene conversion patterns observed at certain human minisatellite loci (Jeffreys *et al.*, 1994).

Whatever the mechanism(s) of concerted evolution, the evolutionary stability of the primate U2 array cannot be explained by selection for high levels of U2 gene expression because the multiple U2 genes are dispersed in the galago (a prosimian) and in rodents (Dahlberg and Lund, 1988). We therefore favor the most parsimonious hypothesis that the U2 tandem repeat, once established in the primate lineage, contained sequence elements favoring both the stability and the concerted evolution of the array *in situ*. The fact that both junctions between the U2 tandem array and flanking chromosomal DNA lie within the solo LTR (Figure 1) suggests that the LTR itself may be such a recombinogenic sequence element; the powerful U2 transcription unit itself could be another (Ares *et al.*, 1987). Whether transcriptionally inert tandem arrays like alphoid satellite (Warburton *et al.*, 1993) evolve differently from transcriptionally active tandem arrays like the primate *RNU2* locus remains to be seen.

Paradoxically, the evolutionary stability of the *RNU2* locus requires constant change: mutations arising in any copy of the U2 repeat unit (including the 5 kb proviral deletion) must be purged from that copy or spread to every other copy (genetically fixed). An interesting question, therefore, is whether continual remodeling of *RNU2* might be mechanistically related to pathological or evolutionary chromosome instability. In fact, infection of human cells with highly oncogenic adenovirus type 12 is known to generate four major chromosome fragile sites (also known as modification sites). These four sites map to the U1 snRNA genes and pseudogenes at 1p36 and 1q21 respectively (Lindgren *et al.*, 1985b), the U2 snRNA genes at 17q21 (Lindgren *et al.*, 1985a) and the 5S rRNA genes at 1q42–43 (Sorensen *et al.*, 1991). The implication may be that a high local concentration of efficient transcription units, or some other aspect of tandemly repeated DNA sequence organization, is responsible for the specificity of virally induced chromosome fragility. Remarkably, all four of these adenovirus 12 inducible fragile sites also colocalize with common fragile sites that are sometimes rearranged in human tumors (Schramayr *et al.*, 1990; Caporossi *et al.*, 1991). Since a majority of the breakpoints observed in the evolution of primate chromosomes occur at or near known fragile sites (Miro *et al.*, 1987; Smeets and van de Klundert, 1990), the cytological stability of *RNU2* becomes all the more puzzling in light of these data.

The U2 repeat unit contains a nearly pure $(CT)_{70}$ tract (position 5084–5222) flanked by CT-rich sequences. The existence of extensive CT tracts in at least two other repeat units—the human U1 snRNA genes (Bernstein *et al.*, 1985; Dahlberg and Lund, 1988) and the histone genes of the sea urchin *Psammechinus miliaris* (Hentschel, 1982)—suggests that such tracts may be a cause or consequence of the molecular mechanism(s) responsible for concerted evolution. Alternating copolymers are known to stimulate recombination between extrachromosomal substrates in transfected mammalian cells (Stringer, 1985) and it has been suggested that the ability of CT tracts to assume unusual triplex structures may play a role in concerted evolution of human U1 (Htun *et al.*, 1984) and U2 genes (Htun *et al.*, 1985). Comparison of our CT tract sequence with that obtained from another individual (Htun *et al.*, 1985) revealed occasional dinucleotide insertions and deletions (D.Liao and A.M.Weiner, unpublished data). This is consistent with the observation that simple repetitive sequences are often highly polymorphic in eukaryotes (Armour and Jeffreys, 1992), presumably reflecting the difficulty of repairing (Strand *et al.*, 1993) and replicating (Samadashwily *et al.*, 1993) such unusual DNA sequences. The CT tract in the U2 repeat unit, however, is a repeat within a repeat; this may enable us to estimate the rate of concerted evolution by determining whether CT tracts within individual U2 arrays accumulate sequence heterogeneity faster than it is eliminated by homogenization of the entire array.

Taken together, our data lay the groundwork for two complementary kinds of studies. We can now investigate the rate and boundaries of concerted evolution at *RNU2* by using a *Sac*I polymorphism within the U2 repeat unit to follow the inheritance of individual U2 arrays through three generations of a large Amish kindred. Our current data indicate that U2 tandem arrays are rapidly homogenized after reciprocal recombination between parental alleles, or that reciprocal recombination never occurs (T.Pavelitz, J.Kidd, K.Kidd and A.M.Weiner, unpublished data). We can also ask whether concerted evolution, chromosome fragility and genomic instability are mechanistically related. The ability of an ectopic U2 array to generate a novel adenovirus 12 inducible chromosome fragile site (Li *et al.*, 1993) implies that the U2 snRNA transcription unit, the LTR, the CT tract, or some other feature of the U2 repeat unit interferes with DNA replication and/or chromatin condensation. We are currently determining which U2 repeat sequences are required to generate a virally inducible fragile site and whether these same sequences can influence amplification of a linked selectable marker, as might be expected if the current U2 array were generated or maintained by natural gene amplification events (Matera *et al.*, 1990; but see Tlsty, 1990; Livingstone *et al.*, 1992).

Materials and methods

In situ hybridization

Chromosome preparations, DAPI staining, labeling of probes, suppression of repetitive sequence signals, chromosome painting and digital

imaging microscopy were essentially as described (Ried *et al.*, 1992). The chromosome 17 paint was generated using a chromosome-specific plasmid library (Collins *et al.*, 1991; gift of J.Gray, UCSF). The *RNU2* probes were pTP18 (a complete *Hin*dIII U2 repeat unit cloned into pUC118) and JL11 described below. The *NGFR* probe was 30 kb cosmid clone (gift of H.Ogura and K.Kidd). Orang utan, gorilla and chimpanzee lymphocyte lines immortalized with Epstein–Barr virus were a gift of D.Lawlor (Stanford University) and were established from blood samples provided by O.Ryder (San Diego Zoo). Baboon blood was obtained from B.Innis (Yale University).

Cloning and analysis of repeat unit and junction fragments

HT1080 cells, a pseudodiploid human fibrosarcoma with a modal chromosome number of 46 (Li *et al.*, 1993), were embedded in agarose plugs (InCert) and prepared for field inversion gel electrophoresis (FIGE) as described (Ausubel *et al.*, 1990). Digestion of the plugs with *Eco*RI, which does not cut the human U2 repeat unit (Van Arsdell and Weiner, 1984a), yielded intact arrays which were well resolved by room temperature FIGE through 0.8% low melting agarose (SeaKem) using 0.2×TBE buffer, 0.5 μg/ml ethidium bromide, a field strength of 15 V/cm, and initial forward and reverse times of 0.3 and 0.1 s increased by 90 increments of 0.03 and 0.01 s respectively. Intact U2 arrays were localized relative to high molecular weight markers (λ concatemers, New England Biolabs) and the localization confirmed by blotting fractions from the preparative gel. Junction fragments were derived from the intact arrays by melting the gel slice at 65°C and redigesting with *Hin*dIII. The resulting *Hin*dIII–*Eco*RI junction fragments were ligated between the *Eco*RI and *Hin*dIII sites of pUC118, and introduced into *E. coli* DH10 by electroporation (Bio-Rad Gene Pulser). Two thousand colonies were gridded and probed with a 3.7 kb *Pvu*II–*Hin*dIII fragment from the right end of the U2 repeat unit (Figure 1): four colonies proved positive and remained positive on rescreening. Clones 7 and 9 appeared identical, clone 17 had contained a 150 bp insertion relative to clones 7 and 9, and clone 11 also reacted with a 584 bp *Hin*dIII–*Pvu*II fragment from the left end of the U2 repeat unit. A complete 6 kb *Hin*dIII U2 repeat unit was cloned into pUC119 from the DNA of Dr W.-j.Poo (Department of Medicine, Yale School of Medicine). Subcloned fragments were sequenced directly by the enzymatic method or, when necessary, after generating nested sets of unidirectional deletions (Ausubel *et al.*, 1990). More than 70% of the sequence was confirmed on both strands. The complete sequence was arbitrarily numbered from the unique *Hin*dIII site and deposited in GenBank (accession No. L37793). Sequence analysis was performed using the Wisconsin Genetics Computer Group suite of programs; all details of the analysis presented as data not shown are available upon request.

Acknowledgements

We thank Asher Davidson and Jakub Buchowski for contributions to this project, members of our laboratory for comments on the manuscript, Drs Judith and Kenneth Kidd for cells and consulation, and Dr David C.Ward for use of imaging facilities. A.G.M. was supported by Post-doctoral Fellowship DRG-1135 from the Damon Runyon–Walter Winchell Cancer Fund. This work was supported by NIH Awards GM41624 and GM31073 to A.M.W.

References

Amstutz,H., Munz,P., Heyer,W.D., Leupold,U. and Kohli,J. (1985) *Cell*, **40**, 879–886.

Ares,M.,Jr, Chung,J.-S., Giglio,L. and Weiner,A. M. (1987) *Genes Dev.*, **1**, 808–817.

Armour,J.A. and Jeffreys,A.J. (1992) *Curr. Opin. Genet. Dev.*, **2**, 850–856.

Armour,J.A., Wong,Z., Wilson,V., Royle,N.J. and Jeffreys,A.J. (1989) *Nucleic Acids Res.*, **17**, 4925–4935.

Armour,J.A., Harris,P.C. and Jeffreys,A.J. (1993) *Hum. Mol. Genet.*, **2**, 1137–1145.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1990) *Current Protocols In Molecular Biology*. John Wiley, New York.

Bernstein, L.B., Manser,T. and Weiner,A.M. (1985) *Mol. Cell. Biol.*, **5**, 2159–2171.

Bollag,R.J., Elwood,D.R., Tobin,E.D., Godwin,A.R. and Liskay,R.M. (1992) *Mol. Cell. Biol.*, **12**, 1546–1552.

Buetti,E. (1994) *Mol. Cell. Biol.*, **14**, 1191–1203.

Caporossi,D., Bacchetti,S. and Nicoletti,B. (1991) *Cancer Genet. Cytogenet.*, **54**, 39–53.

Collins,C., Kuo,W.L., Segraves,R., Fuscoe,J., Pinkel,D. and Gray,J. (1991) *Genomics*, **11**, 997–1006.

Curtis,D., Clark,S.H., Chovnick,A. and Bender,W. (1989) *Genetics*, **122**, 653–661.

Dahlberg,J.E. and Lund,E. (1988) In Birnstiel,M. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer Verlag, Heidelberg, pp. 38–70.

Djian,P. and Green,H. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 8447–8451.

Dombroski,B.A., Mathias,S.L., Nanthakumar,E., Scott,A.F. and Kazazian,H.H.,Jr (1991) *Science*, **254**, 1805–1808.

Doolittle,R.F. and Feng,D.F. (1992) *Curr. Topics Microbiol. Immunol.*, **176**, 195–211.

Dover,G.A. (1993) *Curr. Opin. Gen. Dev.*, **3**, 902–910.

Durnam,D.M., Menninger,J.C., Chandler,S.H., Smith,P.P. and McDougall,J.K. (1988) *Mol. Cell. Biol.*, **8**, 1863–1867.

Edelman,G.M. and Gally,J.A. (1970) In Schmitt,F.O. (ed.), *Neurosciences: Second Study Program*. Rockefeller University Press, New York, 962pp.

Eikenboom,J.C.J., Vink,T., Briet,E., Sixma,J.J. and Reitsma,P.H. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 2221–2224.

Elder,J.F. and Turner,B.J. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 994–998.

Fink,G.R., Boeke,J.D.and Garfinkel,D.J. (1986) *Trends Genet.*, **2**, 118–123.

Gonzalez, I.L., Wu,S., Li,W.M., Kuo,B.A. and Sylvester,J.E. (1992) *Nucleic Acids Res.*, **20**, 5846.

Gray,I.C. and Jeffreys,A.J. (1991) *Proc. R. Soc. Lond., Ser. B: Biol. Sci.*, **243**, 241–253.

Hammarstrom,K., Westin,G., Bark,C., Zabielski,J. and Pettersson,U. (1984) *J. Mol. Biol.*, **179**, 157–169.

Hammarstrom,K., Santesson,B., Westin,G. and Pettersson,U. (1985) *Exp. Cell Res.*, **159**, 473–478.

Hellmann-Blumberg,U., Hintz,M.F., Gatewood,J.M. and Schmid,C.W. (1993) *Mol. Cell. Biol.*, **13**, 4523–4530.

Hentschel,C.C. (1982) *Nature*, **295**, 714–716.

Hipeau-Jacquotte,R., Brutlag,D.L. and Bregegere,F. (1989) *Mol. Gen. Genet.*, **220**, 140–146.

Htun,H., Lund,E. and Dahlberg,J.E. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 7288–7292.

Htun,H., Lund,E., Westin,G., Pettersson,U. and Dahlberg,J.E. (1985) *EMBO J.*, **4**, 1839–1845.

Jackson,M.S., Mole,S.E. and Ponder,B.A.J. (1992) *Nucleic Acids Res.*, **20**, 4781–4787.

Jeffreys,A.J., Tamaki,K., MacLeod,A., Monckton,D.G., Neil,D.L. and Armour,J.A. (1994) *Nature Genet.*, **6**, 136–145.

Jinks-Robertson,S. and Petes,T.D. (1993) *Methods Enzymol.*, **224**, 631–646.

Li,Y.P., Tomanin,R., Smiley,J.R. and Bacchetti,S. (1993) *Mol. Cell. Biol.*, **13**, 6064–6070.

Lindgren,V., Ares,M., Weiner,A.M. and Francke,U. (1985a) *Nature*, **314**, 115–116.

Lindgren,V.L., Bernstein,L.B., Weiner,A.M. and Francke,U. (1985b) *Mol. Cell. Biol.*, **5**, 2172–2180.

Livingstone,L.R., White,A., Sprouse,J., Livanos,E., Jacks,T. and Tlsty,T.D. (1992) *Cell*, **70**, 923–935.

McCombie,W.R. *et al.* (1992) *Nature Genet.*, **1**, 348–353.

Martignetti,J.A. and Brosius,J. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 11563–11567.

Matera,A.G. and Marks,J. (1993) *J. Hum. Evol.*, **24**, 233–238.

Matera,A.G., Weiner,A.M. and Schmid,C. (1990) *Mol. Cell. Biol.*, **10**, 5876–5882.

Miro,R., Clemente,I.C., Fuster,C. and Egozcue,J. (1987) *Hum. Genet.*, **75**, 345–349.

Murti J.R., Bumbulis,M. and Schimenti,J.C. (1992) *Mol. Cell. Biol.*, **12**, 2545–2552.

Nag,D.K. and Petes,T.D. (1990) *Mol. Cell. Biol.*, **10**, 4420–4423.

Ono,M., Yasunaga,T., Miyata,T. and Ushikubo,H. (1986) *J. Virol.*, **60**, 589–598.

Orr-Weaver,T.L. and Szostak,J.W. (1985) *Microbiol. Rev.*, **49**, 33–58.

Pesce,C.G., Rossi,M.S., Muro,A.F., Reig,O.A., Zorsopulous,J. and Kornblihtt,A.R. (1994) *Nucleic Acids Res.*, **22**, 656–661.

Ried,T., Baldini,A., Rand,T.C. and Ward,D.C. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1388–1392.

Samadashwily,G.M., Dayn,A.and Mirkin,S.M. (1993) *EMBO J.*, **12**, 4975–4984.

Schmid,C. and Maraia,R. (1992) *Curr. Opin. Genet. Dev.*, **2**, 874–882.
Schramayr,S., Caporossi,D., Mak,I., Jelinek,T. and Bacchetti,S. (1990) *J. Virol.*, **64**, 2090–2095.
Seperack,P., Slatkin,M. and Arnheim,N. (1988) *Genetics*, **119**, 943–949.
Smeets,D.F.C.M. and van de Klundert,F.A.J.M. (1990) *Cytogenet. Cell Genet.*, **53**, 8–14.
Smith,G.P. (1974a) *Science*, **191**, 528–535.
Smith,G.P. (1974b) *Cold Spring Harbor Symp. Quant. Biol.*, **38**, 507–513.
Sorensen,P.D. and Frederiksen,S. (1991) *Nucleic Acids Res.*, **19**, 4147–4151.
Sorensen,P.D., Lomholt,B., Frederiksen,S. and Tommerup,N. (1991) *Cytogenet. Cell Genet.*, **57**, 26–29.
Strand,M., Prolla,T.A., Liskay,R.M. and Petes,T.D. (1993) *Nature*, **365**, 274–276.
Stringer,J.R. (1985) *Mol. Cell. Biol.*, **5**, 1247–1259.
Sweetser,D.B., Hough,H., Whelden,J.F., Arbuckle,M. and Nickoloff,J.A. (1994) *Mol. Cell. Biol.*, **14**, 3863–3875.
Thompson,C.B. (1992) *Trends Genet.*, **8**, 416–422.
Tlsty,T.D. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 3132–3136.
Trask,B.J. and Hamlin,J.L. (1989) *Genes Dev.*, **3**, 1913–1925.
Truss,M., Chalepakis,G. and Beato,M. (1992) *J. Steroid Biochem. Mol. Biol.*, **43**, 365–378.
Van Arsdell,S.W. and Weiner,A.M. (1984) *Mol. Cell. Biol.*, **4**, 492–499.
Van Arsdell,S.W. and Weiner,A.M. (1984) *Nucleic Acids Res.*, **12**, 1463–1471.
Warburton,P.E., Waye,J.S. and Willard,H.F. (1993) *Mol. Cell. Biol.*, **13**, 6520–6529.
Weiner,A.M. and Denison,R.A. (1983) *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1141–1149.
Westin,G., Zabielski,J., Hammarstrom,K., Monstein,H.J., Bark,C. and Pettersson,U. (1984) *Proc. Natl Acad. Sci. USA*, **81**, 3811–3815.
Wevrick,R., Willard,V.P. and Willard,H.F. (1992) *Genomics*, **14**, 912–923.
Wienberg,J., Jauch,A., Stanyon,R. and Cremer,T. (1990) *Genomics*, **8**, 347–350.
Wu,T.-C. and Lichten,M. (1993) *Science*, **263**, 515–518.
Yunis,J.J. and Prakash,O. (1982) *Science*, **215**, 1525–1530.

Received on June 10, 1994; revised on September 14, 1994