Concerted evolution of the tandem array encoding primate U2 snRNA (the *RNU2* locus) is accompanied by dramatic remodeling of the junctions with flanking chromosomal sequences

Thomas Pavelitz¹, Daiqing Liao^{1,2} and Alan M. Weiner^{1,3,4}

Departments of ¹Molecular Biophysics and Biochemistry and ⁴Genetics, Yale University School of Medicine, New Haven, CT 06510-8024, USA

²Present address: Département de Microbiologie et d'Infectiologie, Université de Sherbrooke, Faculté de Médecine, 3001, 12^e Avenue Nord, Sherbrooke, Québec, Canada J1H 5N4

³Corresponding author
e-mail: weiner@biomed.med.yale.edu

The genes encoding primate U2 snRNA are organized as a nearly perfect tandem array (the *RNU2* locus) that has been evolving concertedly for >35 Myr since the divergence of baboons and humans. Thus the repeat units of the tandem array are essentially identical within each species, but differ between species. Homogeneity is maintained because any change in one repeat unit is purged from the array or fixed in all other repeats. Intriguingly, the cytological location of *RNU2* has remained unchanged despite concerted evolution of the tandem array. We had found previously that junction sequences between the U2 tandem array and flanking DNA were subject to remodeling over a region of 200–300 bp during the past 5 Myr in the hominid lineage. Here we show that the junctions between the U2 tandem array and flanking DNA have undergone dramatic rearrangements over a region of 1 to >10 kbp in the 35 Myr since divergence of the Old World Monkey and hominid lineages. We argue that these rearrangements reflect the high level of genetic activity required to sustain concerted evolution, and propose a model to explain why maintenance of homogeneity within a tandemly repeated multigene family would lead to junctional diversity.

Keywords: concerted evolution/gene conversion/*RNU2*/U2 snRNA/unequal sister chromatid exchange

Introduction

Most models for the concerted evolution of tandemly repeated sequences, whether simple dinucleotide repeats, minisatellites, alphoid satellites, rDNA or repeats of arbitrary size and function, have regarded the tandem array as an autonomous genetic element subject solely to internal recombination and selection; the junctions of the tandem array with flanking DNA have been assumed constant, or considered irrelevant, partly out of mathematical necessity and partly because junction sequences often are difficult to identify and clone (for a review see Elder and Turner, 1995; but see Armour *et al.*, 1993; Jeffreys *et al.*, 1994, 1998; May *et al.*, 1996). Indeed, models of sequence maintenance by unequal sister

chromatid exchange (Ohta, 1976; Smith, 1976; Szostak and Wu, 1980; Dover, 1982; Nagylaki and Petes, 1982; Ohta and Dover, 1983; Nagylaki, 1984; Walsh, 1987; Warburton *et al.*, 1993) or pairwise sequence conversion (Dover, 1982; Amstutz *et al.*, 1985; Wolff *et al.*, 1989; Nag and Petes, 1990; Hillis *et al.*, 1991; Jeffreys *et al.*, 1994, 1998; Schlötterer and Tautz, 1994; Rockmill *et al.*, 1995; Gangloff *et al.*, 1996; Buard *et al.*, 1998) are mathematically tractable only if the boundary conditions (i.e. the junction sequences) are considered invariant; recombination events that could remodel junction sequences would introduce a mathematical wild card.

Although theoretical models have shaped our view of how tandem repeats might be maintained, the experimental data—primarily from yeast, *Drosophila* and human rDNA, mammalian minisatellites and primate alphoid satellites—have revealed many unanticipated complications. Specifically, the ratio of unequal sister chromatid exchange to sequence conversion can vary (Wolff *et al.*, 1989), as can the length of the conversion tract (Hillis *et al.*, 1991); the conversion donor can be on a homologous or non-homologous chromosome (Seperack *et al.*, 1988; Jeffreys *et al.*, 1994; Warburton and Willard, 1995); conversion often exhibits polarity, possibly suggesting a role for flanking sequences (Armour *et al.*, 1993; Jeffreys *et al.*, 1994, 1998; May *et al.*, 1996; Buard *et al.*, 1998); sequences can be purged not just by unequal chromatid exchange or gene conversion, but by outright physical destruction through a RAD52-independent exonuclease pathway (Ozenberger and Roeder, 1991); and all of these processes potentially may differ in the mitotic and meiotic phases of the life cycle (Szostak and Wu, 1980; Rockmill *et al.*, 1995; Gangloff *et al.*, 1996). In addition, it is not clear whether the elegant genetic studies in yeast apply to metazoans where recombination appears to be far less proficient, if not actually different in kind (Bollag *et al.*, 1992; Murti *et al.*, 1992). We now argue, based on new data for the primate *RNU2* locus, that the junctions where tandem arrays meet flanking chromosomal DNA cannot be ignored. Rather, junction sequences are likely to represent an informative junkheap of past recombination events, unnecessary for contemporary function but revealing the nature and extent of recombination activity required to maintain a tandem array over evolutionary time.

The U2 snRNA genes of the prosimian galago are dispersed (Matera *et al.*, 1990) like those of rats, mice and hamsters (Dahlberg and Lund, 1988), but primate U2 snRNA genes have been organized as a tandem array for >35 Myr since the divergence of the Old World Monkey and hominoid lineages (Matera *et al.*, 1990). The tandemly repeated primate U2 genes have also been evolving concertedly, i.e. all repeat units within each tandem array (*RNU2* allele) are constrained to be identical or very nearly so; changes as large as a 5 kb deletion (Pavelitz

et al., 1995) or as small as a point mutation (Liao and Weiner, 1995; Liao *et al.*, 1997) are spread rapidly throughout each repeat in the *RNU2* allele. Remarkably, the primate U2 genes (the *RNU2* locus) are evolving *in situ*: the genes have remained at the same chromosomal locus (equivalent to human 17q21) despite multiple speciation events spanning 35 Myr (Pavelitz *et al.*, 1995) and despite the high levels of recombination required to assure concerted evolution (Liao *et al.*, 1997). Thus, paradoxically, the apparent stability of the *RNU2* locus requires constant change: any mutation arising in any copy of the U2 repeat unit must either be purged from that copy or spread to every other copy ('fixed') in the array. Taken together, these observations pose many questions. Why is the U2 tandem repeat an evolutionarily stable form of gene organization when dispersed U2 genes suffice in rodents and prosimians? What are the mechanisms of concerted evolution? Is concerted evolution (and perhaps evolutionary stability) a necessary consequence of tandemly repeated sequence organization, or does concerted evolution require special sequence elements located within each repeat unit and/or flanking the tandem array?

Results and discussion

Comparison of the human and baboon U2 repeat units

The tandemly repeated baboon U2 genes are cut once per 11 kb repeat unit by the 'one cutters' *Bgl*II, *Eco*RI and *Hind*III (Matera *et al.*, 1990; also see Figure 1). A complete 11 kb *Eco*RI U2 repeat unit was cloned readily into the vector Lambda Dash, subcloned into plasmids and sequenced manually by primer walking as described (Pavelitz *et al.*, 1995). To facilitate comparison between the complete baboon and human U2 repeat units, cloned and sequenced as 11 kb *Eco*RI and 6.1 kb *Hind*III fragments, respectively, the sequence of the baboon *Eco*RI fragment was permuted to begin from the equivalent of the human *Hind*III site (ACGCTT in baboon) (Figure 2). The Old World Monkey and human U2 repeat units, although separated by >35 Myr, are remarkably similar except for a homologous recombination (indicated by dotted lines) between the 5' and 3' long terminal repeats (LTRs) of the provirus which generated the solo LTR seen in the hominid lineage (Pavelitz *et al.*, 1995; Liao *et al.*, 1998). The baboon provirus consists of two LTRs, *gag* and *env* sequences; the LTRs belong to the LTR13 family which is well represented by solo LTRs in the human genome (Liao *et al.*, 1998), whereas *gag* and *env* are most closely related to those of the human endogenous retrovirus group HERV-K10 (Pavelitz *et al.*, 1995; Liao *et al.*, 1998; HERV-K10 group reviewed by Mayer *et al.*, 1998).

Importantly, the 5' and 3' LTRs of the baboon provirus (bLTR1 and bLTR2, respectively) exhibit species-specific insertions, deletions and mutations which distinguish them from the human solo LTR (hLTR) (Liao *et al.*, 1998; and data not shown), as well as from each other, as expected if each LTR has been evolving independently since divergence of the baboon and hominid lineages (analysis available upon request). Other features of the primate U2 repeat unit that pre-date this divergence are the two *Alu* insertions just downstream of the provirus (Liao *et al.*,

1997), the CT microsatellite (Liao and Weiner, 1995) and a 3' LINE-1 (L1) fragment just downstream from the U2 coding region (Pavelitz *et al.*, 1995). Overall divergence between the human and baboon U2 repeat units ranges from as little as 3 to >11% over a sliding 60 bp window when the proviral deletion, the hypervariable CT microsatellite, the highly conserved U2 gene and *Alu* insertion (discussed below) are discounted.

Identification and cloning of the baboon *RNU2* flanking sequences

In contrast to the 'one cutters' *Bgl*II, *Eco*RI and *Hind*III (Matera *et al.*, 1990), the 'null cutters' *Hpa*I, *Sac*I and *Xba*I do not cut within the repeat unit and therefore excise the two *RNU2* alleles intact from flanking chromosomal DNA (Figure 1A). Unexpectedly, *Eco*RV is a one cutter for the smaller *RNU2* allele but not the larger (Figure 1A; data not shown). Consistent with previous evidence that changes in any repeat unit are quickly purged or spread throughout the human U2 tandem array (Liao *et al.*, 1997), this polymorphic *Eco*RV site is due to a single C to T mutation at position 3352 in the baboon 11 kb sequence (Figure 2; data not shown). As a control, double digests with *Eco*RV and the null cutters *Hpa*I, *Sac*I or *Xba*I were performed (Figure 1A, right) to rule out the remote possibility that distant flanking *Eco*RV sites might cause the smaller *RNU2* allele to co-migrate fortuitously with the larger. The dark smear at the bottom of the gel corresponds to proviral (Pavelitz *et al.*, 1995), solo LTR (Liao *et al.*, 1998) and U2 pseudogene sequences (Van Arsdell and Weiner, 1984) that react with the probe during low stringency hybridization of this 'unblot' in the absence of formamide (Pavelitz *et al.*, 1995).

To identify the left and right junction fragments (JL and JR) joining the baboon U2 tandem array to flanking chromosomal DNA, genomic DNA was digested with the null cutter *Sac*I, the one cutter *Eco*RI or with both, and probed with a non-repetitive fragment of the U2 repeat unit (Figure 1B, left). In addition to the multicopy 11 kb band corresponding to the U2 repeat unit, digestion with *Eco*RI alone yielded two bands of 4.5 and 15 kb (actually unresolved 15 and 14 kb fragments); double digestion with *Sac*I + *Eco*RI yielded two more bands of 6.5 and 7 kb derived by *Sac*I digestion of the 14 kb *Eco*RI band. As shown below, the 4.5 kb band (indicated by a double asterisk) is a solitary abnormal U2 repeat unit present only in the larger *RNU2* allele; of the remaining single copy bands (indicated by a single asterisk), the 14 and 15 kb *Eco*RI fragments are JL and JR, and the 6.5 and 7 kb *Eco*RI-*Sac*I fragments are JL from the larger and smaller *RNU2* alleles.

To characterize the candidate JL and JR bands further, genomic DNA was digested with the null cutter *Sac*I, and the two *RNU2* alleles were resolved by preparative field inversion gel electrophoresis (FIGE) as in Figure 1A. DNA eluted from successive fractions of the FIGE gel was redigested with *Eco*RI to release JL, JR and monomeric 11 kb U2 repeat units (Figure 1B, lanes 1-7). In addition to a small amount of U2 repeat dimer resulting from incomplete *Eco*RI digestion, both *RNU2* alleles yielded the 15 kb fragment subsequently identified as JR; the larger and smaller *RNU2* alleles yielded 10.6 and 10.9 kb monomeric U2 repeat fragments differing solely

Fig. 1. Initial characterization of intact *RNU2* alleles and flanking sequences from baboon. (A) Resolution of the two *RNU2* loci in baboon previously detected by FISH (Pavelitz *et al.*, 1995). Genomic DNA plugs were digested with the 'null cutters' *HpaI*, *SacI* or *XbaI* which do not cut within the U2 repeat unit, or with *EcoRV* which cuts the smaller *RNU2* allele only (see text). The excised *RNU2* alleles were resolved by field inversion gel electrophoresis (FIGE), dried as an 'unblot' and probed as described previously (Pavelitz *et al.*, 1995) with the baboon 11 kb *EcoRI* U2 repeat unit lacking *Alu* sequences (see Materials and methods). Size markers in kb (left) are MidRange PFG type I Markers (New England Biolabs). (B) Identification of baboon JL and JR. The two *RNU2* alleles were excised from genomic DNA by digestion with the 'null cutter' *SacI*, and resolved by preparative 1% low melting agarose FIGE. The region of the FIGE gel bracketing the two alleles was cut into seven fractions, corresponding to *SacI* fragments of increasing size. Each gel fraction was melted and redigested with *EcoRI* to release the junction fragments JL and JR along with monomeric U2 repeat units. Single asterisks indicate candidates for JL and JR; the double asterisk indicates an abnormal U2 repeat unit from the larger allele (see Figure 2); the identity of the asterisked bands, as determined by further analysis and described in the text, is indicated (right). Size markers in kb (MW, left) are lambda *HindIII* fragments. *SacI*, *EcoRI* and *SacI* + *EcoRI* digests of total genomic DNA are shown for reference. H, *HpaI*; S, *SacI*; RI, *EcoRI*; RV, *EcoRV*; X, *XbaI*.

by an *Alu* element insertion found in all U2 repeats of the smaller array (Figure 2); and the larger and smaller *RNU2* alleles yielded 6.5 and 7.0 kb fragments, respectively, which subsequently were identified as the two JLs which differ solely by insertion of an *Alu* element (Figure 1B, right; also see Figures 2 and 4). Knowing that the candidate JL and JR fragments had *EcoRI* or *EcoRI* + *SacI* termini, we cloned *SacI* + *EcoRI* double digests of genomic DNA between the *EcoRI* and *SacI* sites of Lambda Zap (a mixture of Lambda Zap cut with either *EcoRI* or *SacI*), and also constructed a 12–18 kb size-selected *EcoRI* library in Lambda Dash. The *EcoRI* library yielded the 15 kb JR fragment and a longer version (14 kb) of JL, while the directional *EcoRI*–*SacI* library yielded the 6.5 and 7.0 kb JL fragments and a 4.5 kb abnormal U2 repeat. As expected for authentic junction fragments, non-repetitive probes derived from JL (a 205 bp *SacI*–*TaqI* fragment) and JR (a 242 bp fragment extending leftward from the *EcoRI* site to a *RsaI* site) reacted with both *RNU2* alleles excised by *XbaI* and did not react with the U2 repeat unit (data not shown).

Homogeneity of the baboon arrays

The downstream LTR (bLTR2) of all U2 repeat units from the smaller *RNU2* allele contains an *Alu* insertion (Figure 1B; data not shown). Insertion of this *Alu* element, like

the point mutation which caused gain or loss of the *EcoRV* site (Figures 1A and 2), presumably occurred in a single U2 repeat unit and then spread rapidly to fixation in all other repeats in the same array. Thus concerted evolution of the *RNU2* locus is surprisingly versatile, and can fix point mutations (the *EcoRV* site and the previously studied *SacI* site; Liao *et al.*, 1997), small insertions and deletions (the CT microsatellite; Liao and Weiner, 1995; Liao *et al.*, 1997), large insertions (*Alu* elements) and large deletions (of the HERV-K10-like provirus to generate a solo LTR; Pavelitz *et al.*, 1995).

The abnormal baboon 4.5 kb *EcoRI* U2 repeat unit, which is present as a solitary copy embedded in the larger *RNU2* allele (Figure 1B, double asterisk), must reflect homologous recombination between the 5' and 3' LTRs of two different repeat units because the LTR is flanked on either side by duplicated proviral sequences (Figure 2, bottom two lines). Thus the LTR of the abnormal U2 repeat is neither a 5' nor 3' LTR, but a hybrid between the two (Figure 2, bLTR1/2); moreover, the parental 5' and 3' LTRs are sufficiently divergent that the recombination site can be localized between LTR positions 939 and 958 [data not shown; LTR sequences numbered 1–1007 as in Liao *et al.* (1998)]. Note that this abnormal U2 repeat unit must lie within the larger U2 tandem array, and not external to it, because three different enzymes

Fig. 2. Comparison of the baboon and human U2 repeat units. To facilitate comparison with the previously characterized human 6.1 kb *Hind*III U2 repeat unit (top), the sequence of the cloned baboon 11 kb *Eco*RI U2 repeat unit was permuted and renumbered from the equivalent of the human *Hind*III site [ACGCTT in baboon, designated (H3) here]. Homologous recombination (dotted lines) between the 5' and 3' LTRs of the HERV-K10-like provirus generated a solo LTR in the hominid lineage (Pavelitz *et al.*, 1995; Liao *et al.*, 1998). Also shown is an abnormal baboon 4.5 kb *Eco*RI U2 repeat unit from the larger *RNU2* allele apparently due to homologous recombination between the 5' and 3' LTRs of two different repeat units; the result is an internal duplication of proviral sequences (bottom). All LTRs are ~1 kb in length: hLTR, human solo LTR; bLTR1 and bLTR2, upstream and downstream LTRs of the baboon HERV-K10 provirus; bLTR1/2, hybrid LTR resulting from recombination between bLTR1 and bLTR2. CT, (CT)_{~70} microsatellite; L1, 3' fragment of L1 LINE element; *Alus*, an *Alu*-rich 1700 bp interval containing two complete *Alu* elements (solid arrows) and three 3' *Alu* fragments (solid arrowheads) in the same orientation, followed by an MER33 element (solid box). An *Alu* element and *Eco*RV site found only in the smaller baboon *RNU2* allele are denoted by an asterisk. A, *Ase*I; B, *Bam*HI; D, *Dra*I; Ea, *Eag*I; Ec, *Eco*RI; H2, *Hinc*II; H3, *Hind*III; K, *Kpn*I; Nd, *Nde*I; Ns, *Nsi*I; P, *Pst*I.

(*Eco*RI, *Bam*HI and *Hind*III) release the hybrid LTR as a 4.5 kb fragment of identical size as predicted by the map (Figure 2, bottom; data not shown). This abnormal U2 repeat may have arisen quite recently, as it is the only heterogeneity we have detected within a single allele of the concertedly evolving baboon *RNU2* locus.

The baboon HERV-K10-like provirus persists in sequences immediately flanking the human U2 tandem array

We previously found that the human U2 tandem repeat meets flanking chromosomal DNA within the hLTR at both the left and right junctions (JL and JR) (Pavelitz *et al.*, 1995). Remarkably, human JR turns out, upon further analysis, to be flanked by the 3' end of the baboon HERV-K10-like provirus (bLTR2 and *env*) as shown schematically in Figure 3A (also see Figure 5 below). Not only do species-specific insertions, deletions and mutations clearly identify this flanking LTR as of baboon, not human, origin but the *env* sequences are virtually identical to those of the baboon U2 repeat unit when allowance is made for the expected divergence between baboon and human sequences. Thus, although the baboon HERV-K10-like provirus was deleted from the human U2 tandem repeat unit, leaving behind a solo hLTR, the parental baboon HERV-K10-like provirus persists in sequences immediately flanking the human *RNU2* tandem repeat (Figure 3A). We conclude from the overall sequence organization of JR that: (i) concerted evolution of the

primate *RNU2* locus occurs within the tandem repeat, but does not extend into flanking sequences; (ii) JR has undergone major rearrangements because the junction between the tandem array and flanking sequences has shifted (compare Figures 3 and 4), and the tandem LTRs found in human JR (Figure 3A and B; also see Figure 5 below) are not found in baboon JR; and (iii) sequences flanking the U2 tandem array may provide historical clues regarding the evolution of the tandem repeat, as well as the mechanism(s) of concerted evolution.

The right junction region (JR) has been subject to modest, but repeated remodeling in the hominid lineage

Using the human JL and JR sequences (Pavelitz *et al.*, 1995) to design PCR primers, we were able to amplify JR from chimpanzee and gorilla, but not JL, although human JL was readily amplified in control experiments (see Materials and methods). Thus chimpanzee and gorilla JL appear to have diverged significantly from human; however, a detailed comparison of the human, chimpanzee and gorilla JR sequences (Figure 3B) reinforces our earlier hypothesis (Pavelitz *et al.*, 1995) that the right junction region between the hominid U2 tandem array and flanking chromosomal DNA changes more rapidly than either the U2 repeat unit itself or the flanking sequences.

For this analysis, we defined the junction between hominid *RNU2* and flanking chromosomal sequences as the position where homology breaks off between the

Fig. 3. The human right junction (JR) retains vestiges of baboon *RNU2* sequences, and has been subject to modest remodeling in the hominid lineage. (A) Comparison of the human U2 repeat unit with human JL and JR. The human U2 tandem repeat breaks off within LTR sequences (hLTR) at both JL and JR (Pavelitz *et al.*, 1995). As described in the text, JL is flanked by *Alu*-rich sequences, JR by the 3' end of the baboon HERV-K10-like provirus (bLTR2 and *env*). hLTR sequences are indicated by a black arrow, bLTR sequences by light shading and flanking sequences by dark shading. A unique human sequence motif also found distal to baboon JR is bracketed (see Figure 5 for details). (B) Comparison of human, chimpanzee and gorilla JR (bottom three lines) with the human U2 repeat unit (fourth line from bottom) at higher resolution. Sequences homologous to the U2 tandem repeat unit are indicated by a solid line, flanking sequences by dark shaded boxes and internal 53 bp tandem 'duplicons' by arrows. The genomic structure of hominid JR is shown schematically (top) using the same conventions as in (A). All percentages represent divergence over the indicated sequence interval; key numbers referred to in the text are highlighted. To calculate divergence, mutations and small insertions or deletions (<5 bp) were counted equally, but larger insertion and deletions were discounted to avoid bias. Homology with the U2 repeat unit breaks off sharply at human JL (Pavelitz *et al.*, 1995) and chimpanzee JR, but decreases stepwise at human (Pavelitz *et al.*, 1995) and gorilla JR.

human U2 repeat unit and the three right junction sequences (human, gorilla and chimpanzee), 8 bp from the 5' end of the solo hLTR (Figure 3B). Immediately to the right (telomeric) side of the junction, bLTR2 is found corresponding to the 3' end of the baboon HERV-K10 provirus (Figure 3A and B); immediately to the left (centromeric) side of the junction is the solo LTR of the hominid U2 repeat units (see Materials and methods for the chromosomal orientation of the human *RNU2* locus).

The solo LTR in the junction region contains either two (in gorilla and chimpanzee) or three tandem repeats (in human) of a 53 bp sequence (the 'duplicon') whose copy number also varies in other human solo LTRs of the LTR13 family (Liao *et al.*, 1998). To the left of the second duplicon, human JR agrees best with the human U2 repeat sequence (<1% divergence compared with 5% for chimpanzee JR and 9% for gorilla JR) as expected if the homogeneity of U2 repeat units within the tandem array

is maintained efficiently by concerted evolution. Most surprisingly, chimpanzee JR agrees >4-fold better than does human JR with the human U2 repeat sequence over a 280 bp region between the duplicons and the actual junction site (<5% compared with >20% divergence). In fact, a 28 bp deletion which was present in the solo LTR of the ancestral hominid JR (see discussion of Figure 5 below), and is still found in human and gorilla JR, is missing in chimpanzee JR. The simplest explanation is that the 28 bp deletion in the ancestral chimpanzee JR was repaired by using the complete solo LTR of the chimpanzee U2 repeat unit either as a donor for gene conversion or as a partner for unequal sister chromatid exchange. In any event, we can conclude that concerted evolution locally increases the mutation frequency in the right junction region where U2 tandem repeats meet flanking chromosomal sequences. A mechanism that could explain locally increased mutation frequencies is considered below (see Figure 6).

Although we were unable to clone JL from chimpanzee and gorilla, it is nonetheless intriguing that JL is a sharp boundary in both baboons and humans, whereas JR decays irregularly in hominids, accumulating insertions, deletions and mutations compared with the U2 repeat unit (Figure 3B; detailed sequence analysis available upon request; see also Pavelitz *et al.*, 1995). Thus, although the actual junction regions in baboon and human are completely different (compare Figures 3 and 4) and have been evolving independently for >35 Myr, the baboon and human *RNU2* loci exhibit the same apparent polarity with sharp left and stepwise right boundaries.

The human and baboon junction regions are completely unrelated

The 14 and 15 kb baboon left and right junction fragments, JL and JR, were cloned as described above (Figure 1B; Material and methods). We localized the junction regions by comparing the restriction map of each junction fragment with that of the baboon U2 repeat unit; these regions were then subcloned and sequenced. The JL and JR flanking sequences (shaded boxes in Figure 4) were characterized by a combination of mapping, blotting with a variety of probes and sequencing. In outline, the 6 kb JL flanking sequence was found to contain a HERV-K10 provirus (sequenced rightward from the *EcoRI* site) separated from the junction region (sequenced rightward from *SacI*) by 2 kb of *Alu*-rich DNA with no detectable homology to the baboon U2 repeat unit (*HincII*–*SacI* blotted). The 7.5 kb JR flanking sequence lacked detectable homology to baboon U2 repeat sequences but was *Alu* rich (*SalI*–*EcoRI* blotted) beyond the junction region (sequenced leftward from *HpaII*). Thus, although the baboon and human U2 repeat units differ simply by homologous excision of the baboon HERV-K10 provirus, leaving behind a solo LTR in human (Figure 2), the junction regions of the baboon and human U2 tandem arrays are completely unrelated: the human U2 tandem array meets flanking chromosomal DNA within the LTR (Pavelitz *et al.*, 1995; see also Figure 3A) but the baboon U2 tandem array does not (Figure 4), and there is no obvious homology between baboon and human JL, or baboon and human JR.

Human JR is derived from sequences distal to baboon JL, indicating dramatic rearrangements of *RNU2* flanking sequences

The human U2 tandem repeat is flanked on the right (telomeric) junction by the baboon HERV-K10 provirus, demonstrating that baboon proviral sequences which have been purged from the U2 tandem array still persist in flanking DNA (Figure 3). More remarkably, we found on closer inspection that human JR corresponds to sequences which originally were located distal to baboon JL (see Figure 5, which compares the bracketed regions of Figures 3 and 4). This rearrangement of sequences from the left flank to the right flank of the U2 tandem array cannot be accounted for by any simple inversion of the U2 tandem array between the chromosomal flanks: not only is simple inversion inconsistent with the structure of human (Figure 3) and baboon JL and JR (Figure 4), but the sequences which now form human JR reside within the flanking sequences of baboon JL >2.5 kb away from the junction itself. Thus, although the U2 tandem repeat units within the *RNU2* locus have been evolving in an orderly and concerted fashion since the divergence of baboons and humans, the flanking sequences have undergone dramatic rearrangements.

A detailed analysis of human JR and baboon JL provides additional evidence for these conclusions (Figure 5). In both human JR and baboon JL, tandem LTRs result from juxtaposition of a HERV-K10 provirus immediately upstream of a solo LTR derived from it (Liao *et al.*, 1998). As expected, all four of these LTRs are members of the LTR13 subfamily characteristic of the primate *RNU2* locus (Liao *et al.*, 1998). The upstream LTRs can be identified by diagnostic mutations as bLTR2, and clearly belong to the upstream proviruses. Unexpectedly, the downstream solo LTRs in both human JR and baboon JL exhibit the same unique hallmarks (an 8 bp 5' truncation, and identical 4 and 28 bp internal deletions located close to the 5' end) (Figure 5). As none of 30 other LTR13 family members exhibit these hallmarks (Liao *et al.*, 1998), the strong implication is that these tandem LTRs originally arose in baboon or an ancestor thereof (presumably as a result of proviral insertion immediately upstream of a 5'-truncated, internally deleted solo LTR) and that the resulting DNA was then transposed or rearranged from baboon JL to human JR.

Junctional diversity may be the price of perfection

At least four different mechanisms have been proposed to maintain the homogeneity of the tandemly repeated DNA sequences commonly found in metazoan genomes (Dover, 1993): unequal sister or non-sister chromatid exchange (Smith, 1976), gene conversion (Edelman and Gally, 1970; Jeffreys *et al.*, 1994; Schlötterer and Tautz, 1994; Gangloff *et al.*, 1997), gene amplification (Weiner and Denison, 1983; Matera *et al.*, 1990) and (perhaps only for relatively simple sequence repeat units) replication slippage and/or heteroduplex repair (Armour *et al.*, 1993; Parsons *et al.*, 1993; Strand *et al.*, 1993; Prolla *et al.*, 1994).

All four mechanisms might have been expected to maintain stable boundaries between the U2 tandem array and flanking chromosomal sequences: unequal chromatid exchange because it is a cut-and-paste mechanism; replication slippage and repair because these processes are

Fig. 4. JL and JR have undergone major rearrangements since the divergence of Old World Monkey and hominid lineages. The baboon JL and JR are compared with a complete baboon 11 kb U2 repeat unit. A solo LTR in baboon JL is indicated (bsLTR). *Alu* elements in the flanking sequences were localized to the indicated restriction fragments by blotting with the adjacent *SalI*–*HpaII* fragment which contains a nearly complete *Alu* element; *Alu* copy number was estimated by normalizing the unknown *Alu* signals to that seen for the *SalI*–*HpaII* fragment itself. The *Alu* insertion in bLTR2 is present in all repeat units of the smaller allele; the *Alu* element located near the CT repeat is present only in the left junction of the smaller allele (asterisks). The positions of the null cutters *HpaI*, *SalI*, *SacI* and *XbaI* are indicated; an *EcoRV* site is present only in the smaller allele and the 7 kb JL fragment derived from it. Other conventions are as in Figure 3. A unique baboon sequence motif also found at human JR is bracketed (see Figure 5 for details).

Fig. 5. Human JR is derived from sequences distal to baboon JL. Sequences from human JR (bracketed in Figure 3A) and baboon JL (bracketed in Figure 4) are compared schematically. These tandem LTRs must have arisen in baboon (or an ancestor thereof) because the same 8 bp 5' truncation, and internal 4 and 28 bp deletions, are found in each downstream LTR, but not in any of 30 other members of the LTR13 family. Note that rearrangement of sequences from baboon JL to human JR cannot be explained by a simple inversion of the U2 tandem array between the chromosomal flanks (see text). LTR sequences of baboon origin are lightly shaded; human solo LTR sequences, generated by recombination between baboon LTR1 and LTR2, are solid black. Sequences to the left (centromeric) side of the 28 bp deletion in human JR resemble the human solo LTR more than the baboon solo LTR, consistent with repeated remodeling of human JR (Pavelitz *et al.*, 1995).

templated by existing sequences within the tandem repeat; gene conversion because pairing is usually blocked by even modest heterology (Langin *et al.*, 1988a,b; Iype *et al.*, 1995; Elliott *et al.*, 1998; but see Belmaaza *et al.*, 1994); and gene amplification because the *RNU2* locus has remained at 17q21 for >35 Myr (Pavelitz *et al.*, 1995) thus tending to exclude amplification (and concerted evolution) through reintegration of extrachromosomal intermediates. Contrary to expectation, however, we found that JL and JR have undergone dramatic rearrangements

between the baboon and hominid lineages, and have also shifted in position relative to the U2 repeat unit: the U2 tandem array meets flanking chromosomal sequences within the solo LTR in human JL and JR (Figure 3A) but far from the LTRs in both baboon JL and JR (Figure 4).

Our data documenting dramatic rearrangements of *RNU2* junction sequences argue strongly against unequal chromatid exchange as the mechanism of concerted evolution, at least for the primate *RNU2* locus, because simple reciprocal exchanges between homologous sequences cannot readily explain movement of the junctions themselves or dramatic rearrangements of the sequences flanking them (Ohta, 1976; Smith, 1976; Szostak and Wu, 1980; Dover, 1982; Nagylaki and Petes, 1982; Ohta and Dover, 1983; Nagylaki, 1984; Walsh, 1987; Warburton *et al.*, 1993). Although replication slippage remains a formal explanation for concerted evolution of large repeat units, the strongest experimental evidence clearly favors some form of gene conversion as the predominant mechanism (Gangloff *et al.*, 1997).

Does concerted evolution of the tandem array cause dramatic rearrangements of flanking sequences?

To explain why both JL and JR have undergone dramatic rearrangements on an evolutionary time scale, we propose that the mechanism of concerted evolution causes recombination intermediates to accumulate at the boundaries of the U2 tandem array, and that these intermediates are recombinogenic. Specifically, we propose that the high local concentration of tandem repeats within the *RNU2* array and the ability of these tandem repeats to align in multiple registers favor extensive meiotic and/or mitotic pairing between *RNU2* arrays on homologous chromosomes. This would be consistent with the high frequency of gene conversion documented for the natural tandem array of rRNA genes in *Saccharomyces cerevisiae* (Gangloff *et al.*, 1997) and for an artificial tandem array of *Adh* genes in *Drosophila* (Hipeau-Jacquotte *et al.*,

Fig. 6. A model to explain how concerted evolution of a tandem array could generate junctional diversity and rearrangements. Extensive alignment of homologous *RNU2* loci would allow a single gene conversion event to homogenize multiple U2 repeat units; however, the inability of terminal U2 repeat units to pair with flanking sequences might block resolution of recombination intermediates, leading to localized error-prone repair and/or junctional rearrangements.

1989). Multiple alignment would enable a single gene conversion event to homogenize multiple tandem repeats, just as pairwise alignment of individual genes has been shown experimentally to maintain dispersed multigene families in the yeasts *S.cerevisiae* (Nag and Petes, 1990) and *Schizosaccharomyces pombe* (Amstutz *et al.*, 1985). However, homologous tandem arrays need not align in perfect register (Szostak and Wu, 1980) and the terminal repeats of each array will sometimes be 'paired' with non-homologous flanking sequences (Figure 6). Assuming (as seems likely) that gene conversion initiates within the paired region of the tandem array and then moves outward toward one or both junctions, recombination intermediates would be blocked by sequence heterology at the junctions, and junctional rearrangements might ensue occasionally if the blocked recombination intermediates were difficult to resolve.

This mechanism could account for remodeling of the junction sequences on a small scale during successive rounds of concerted evolution (Figure 3B) as well as for occasional rearrangements of junction sequences on a larger scale (Figures 3A, 4 and 5). If blocked recombination intermediates were repaired by a mutagenic polymerase, as observed for repair of HO-induced double-strand breaks by DNA polymerase ζ in *S.cerevisiae* (Holbeck and Strathern, 1997; for a review of mutagenic repair, see Harris *et al.*, 1999), the model could also account for the high frequency of mutations observed in the immediate vicinity of JR (Figure 3B). In addition, the model could explain why neither the baboon nor human *RNU2* loci span an integral number of U2 repeat units, or terminate at a fixed site within the repeat unit (Figures 3 and 4): junctions between the tandem array and flanking chromosomal DNA would not represent obligatory initiation or termination sites for recombination, but rather arbitrary sites where outwardly moving recombination intermediates happened to be resolved. Finally, the model could even account for preferential remodeling of human JR (Figure 3B and legend) if recombination complexes initiating within the U2 tandem array were polar. Although the conversion gradients observed in mammalian minisatellite arrays have been interpreted as

suggesting a role for flanking sequences (Armour *et al.*, 1993; Jeffreys *et al.*, 1994, 1998; May *et al.*, 1996; Buard *et al.*, 1998), it also seems possible that such gradients could be caused by polar recombination complexes initiating within minisatellite sequences.

Support for key steps in the model comes from studies of 'one-sided homologous recombination' events observed during chromosomal integration of linear DNA (Ellis and Bernstein, 1989; Berinstein *et al.*, 1992; Richard *et al.*, 1994, 1997; Villemure *et al.*, 1997). When exogenous linear DNA containing one homologous and one non-homologous end invades chromosomal DNA, invasion initiates cleanly at the homologous end; however, resolution of the resulting recombination intermediates at the non-homologous end involves illegitimate recombination, and is often accompanied by significant rearrangements with nearby (not necessarily adjacent) chromosomal sequences (Dellaire *et al.*, 1997). Moreover, mild DNA sequence divergence ('homeologous recombination') does not affect initiation of one-sided recombination events, but favors imprecise resolution (Belmaaza *et al.*, 1994).

The similarity to our model is striking. Out-of-register alignment of the two tandem repeats would in effect create substrates for one-sided homologous recombination at either junction; resolution of the recombination intermediates would be unaffected by mild sequence divergence, leading to remodeling of the junctions (Figure 3B; Pavelitz *et al.*, 1995), but greater divergence could block remodeling and lead to major rearrangements (Figure 5, also compare Figures 3 and 4). Consistent with this scenario, one function of the RuvAB proteins in *Escherichia coli* may be to avoid deleterious inversions or deletions by enabling recombination complexes that are blocked by heterology to reverse direction and disengage (Iype *et al.*, 1995; Seigneur *et al.*, 1998).

A key assumption of our model is that heterology can block resolution of recombination intermediates as they approach the boundaries of the array. This cannot invariably be the case, however, or large insertions (*Alu* elements) and deletions (homologous proviral excision) which arise initially in only one U2 repeat unit could never be fixed in all repeats of a single U2 tandem array. We can imagine two explanations, both consistent with our model, which are not mutually exclusive. First, heterology may reduce but not entirely abolish faithful resolution of stalled recombination intermediates. Secondly, gene conversion may more easily traverse heterology sandwiched between homologous U2 repeat units (like the initial *Alu* insertion or proviral deletion) than heterology located at the junction between the U2 tandem array and non-homologous flanking sequences. Assuming that resolution of recombination intermediates spanning an insertion or deletion is mechanistically related to gap repair, both explanations are supported by work showing that high sequence homology favors but is not absolutely required for precise gap repair (Belmaaza *et al.*, 1994).

Perspective

Concerted evolution of the *RNU2* tandem array creates the illusion of order and stability, because all repeat units within the tandem array are constrained to be identical or nearly so. In reality, however, the Red Queen is running fast to stay in place (Van Valen, 1973): the illusion of

genomic stability requires high levels of recombination as revealed by our analysis of the tandem array itself (Liao *et al.*, 1997) and the *RNU2* junction regions (this study). Maintenance of a tandemly repeated multigene family is neither as effortless, nor as free of genetic risk, as might appear; genomic instability is simply pushed outward to the edges of the array.

Materials and methods

Genomic DNAs

Genomic DNAs were from the Epstein–Barr virus (EBV)-transformed lymphocyte lines of human, chimpanzee (*Pan troglodytes*, *Pan paniscus*), gorilla (*Gorilla gorilla*) and baboon (*Pan anubis*). All primate lines were the kind gift of J. and K. Kidd. JR sequences obtained from two different individuals of *P. troglodytes* and one of *P. paniscus* differed only trivially for the purposes of this analysis.

Sequence analysis

All sequence analysis was performed using the Baylor College of Medicine Search Launcher (<http://gc.bcm.tmc.edu:8088/search-launcher.html>). Repetitive sequence elements were located and identified by RepeatMasker. Sequences reported in this paper have been deposited in the DDBJ/EMBL/GenBank database (accession Nos AF147266–AF147277). All details of the sequence analysis are available upon request.

Cloning the baboon U2 repeat unit and flanking sequences

The baboon 11 kb *EcoRI* U2 repeat unit was cloned into the *EcoRI* site of Lambda Dash as described (Matera *et al.*, 1990; Pavelitz *et al.*, 1995). To clone potential junction fragments, we first enriched ~15- to 20-fold for intact *RNU2* arrays by digesting genomic DNA plugs with the 'null cutter' *SacI*, and isolating high molecular weight *SacI* DNA fragments (>100 kb) that run at the exclusion limit of an 0.8% low melting agarose gel under standard FIGE conditions. Gel fractions containing the large *SacI* fragments were melted, digested simultaneously with β -agarase and *EcoRI* (New England Biolabs) and cloned between the *EcoRI* and *SacI* sites of Lambda Zap (Stratagene). The directional *EcoRI*–*SacI* lambda library was screened using the entire 11 kb baboon U2 repeat unit, except for the *Alu* sequence, as probe; this was prepared by random hexamer labeling of three restriction fragments (*EcoRI*–*NsiI*, *NsiI*–*KpnI* and *ApaI*–*EcoRI*). Screening yielded the 6.5 and 7.0 kb *SacI*–*EcoRI* JL fragments from the large and small arrays respectively, and the abnormal 4.5 kb *EcoRI* U2 repeat unit from the larger array. FIGE conditions used for human *RNU2* alleles (Pavelitz *et al.*, 1995) were modified to separate the larger baboon arrays (MJ Research Power Inverter, parameters A, 0.5; B, 0.1; C, 1.5; D, 0.3; E, 90; F, 0; G, 0). An *EcoRI* lambda library was prepared by fractionating *EcoRI*-digested genomic DNA by electrophoresis on a 0.8% low melting agarose gel; 12–18 kb *EcoRI* fragments were eluted by the LiCl method (Farve, 1992), recovered using oyster glycogen as carrier (Tracy, 1981), and cloned directly into the *EcoRI* arms of Lambda Dash. Screening of the *EcoRI* library with the same U2 probe yielded the 15 kb JR fragment and the longer forms of JL as 14 kb *EcoRI* fragments; size selection for >12 kb fragments reduced the yield of 11 kb *EcoRI* U2 repeat units to tolerable levels. Agarose gels and colony lifts were transferred to nylon membranes (CUNO Zetabind) under alkaline conditions, and probed at high stringency as described to avoid reaction with HERV-K10 proviral sequences (Pavelitz *et al.*, 1995), solo LTR13 elements (Liao *et al.*, 1998) and U2 pseudogenes (Van Arsdell and Weiner, 1984).

Cloning JR of great apes

JR was amplified from human, chimpanzee and gorilla DNA by PCR using the same primers U2JR1 (5'-ACCACTGAAGCACAGCATCA-3') and U2JR2 (5'-TAACAGCGTAGCTAGCCTTC-3') and conditions previously described (Liao *et al.*, 1997). The resulting PCR fragments were gel purified, cloned into pGEM-T (Promega) and propagated in *E. coli* DH10B α as described previously (Liao and Weiner, 1995). Optimal PCR conditions were determined by increasing the MgCl₂ concentration from 0.5 to 4 mM in 0.5 mM increments, or the annealing temperature in 2°C increments. If these measures failed to generate a visible fragment, the desired PCR product was detected by blotting, eluted, and reamplified with the same primers. Although primers U2JL1 (5'-AGACTGAGGCATGAGAATCA-3') and U2JL2 (5'-ACACAGAGTTAGGAGCTGAA-3')

readily amplified the expected 595 bp human JL fragment using genomic DNA as template (Liao *et al.*, 1997), we were unable to amplify gorilla and chimpanzee JL from total genomic DNA or from isolated primate *RNU2* arrays using these or numerous other primer pairs. Conceivably, human JL has diverged significantly from that of other primates, although JR has not.

Chromosomal orientation of *RNU2* arrays

To determine the orientation of the human *RNU2* array at 17q21, we performed PCR using primer pairs specific for human JL (U2JL1 and U2JL2) and JR (U2JR1 and U2JR2), and an ordered array of four overlapping P1 genomic clones extending toward the telomere from *BRCA1* (Neuhausen *et al.*, 1994) as template. These four clones (#402, 1046, 405 and 931; kind gift of Russell Bell, Myriad Genetics) in the P1 vector pAd10SacBII (Pierce *et al.*, 1992) were located between markers *D17S1327* and *D17S1325*. JL could be amplified from clone #405, JR from #931 (data not shown).

Acknowledgements

We thank A.D. Bailey for excellent advice throughout this project and acute comments on the manuscript, J. and K. Kidd for generously providing primate cell lines, P. Chartrand for encouragement, and all three reviewers for helpful criticism. This work was supported by a Postdoctoral Fellowship from the Medical Research Council of Canada (D.L.) and National Institutes of Health Award GM41624 (A.M.W.).

References

- Amstutz, H., Munz, P., Heyer, W.D., Leupold, U. and Kohli, J. (1985) Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell*, **40**, 879–886.
- Armour, J.A.L., Harris, P.C. and Jeffreys, A.J. (1993) Allelic diversity at minisatellite MS205 (*D16S309*): evidence for polarized variability. *Hum. Mol. Genet.*, **2**, 1137–1145.
- Bailey, A.D., Li, Z., Pavelitz, T. and Weiner, A.M. (1995) Adenovirus 12 induced fragility of the human *RNU2* locus requires U2 snRNA transcriptional regulatory elements. *Mol. Cell. Biol.*, **15**, 6246–6255.
- Belmaaza, A., Milot, E., Villemure, J.F. and Chartrand, P. (1994) Interference of DNA sequence divergence with precise recombinational DNA repair in mammalian cells. *EMBO J.*, **13**, 5355–5360.
- Berinstein, N., Pennell, N., Ottaway, C.A. and Shulman, M.J. (1992) Gene replacement with one-sided homologous recombination. *Mol. Cell. Biol.*, **12**, 360–367.
- Bollag, R.J., Elwood, D.R., Tobin, E.D., Godwin, A.R. and Liskay, R.M. (1992) Formation of heteroduplex DNA during mammalian intrachromosomal gene conversion. *Mol. Cell. Biol.*, **12**, 1546–1552.
- Buard, J., Bourdet, A., Yardley, J., Dubrova, Y. and Jeffreys, A.J. (1998) Influences of array size and homogeneity on minisatellite mutation. *EMBO J.*, **17**, 3495–3502.
- Dahlborg, J.E. and Lund, E. (1988) The genes and transcription of the major small nuclear RNAs. In Birnstiel, M. (ed.), *Structure and Function of Major and Minor Small Nuclear Ribonucleoprotein Particles*. Springer Verlag, Heidelberg, Germany, pp. 38–70.
- Dellaire, G., Lemieux, N., Belmaaza, A. and Chartrand, P. (1997) Ectopic gene targeting exhibits a bimodal distribution of integration in murine cells, indicating that both intra- and interchromosomal sites are accessible to the targeting vector. *Mol. Cell. Biol.*, **17**, 5571–5580.
- Dover, G.A. (1982) Molecular drive, a cohesive mode of species evolution. *Nature*, **299**, 111–117.
- Dover, G.A. (1993) Evolution of genetic redundancy for advanced players. *Curr. Opin. Genet. Dev.*, **3**, 902–910.
- Edelman, G.M. and Gally, J.A. (1970) Arrangement and evolution of eukaryotic genes. In Schmitt, F.O. (ed.), *Neurosciences: Second Study Program*. Rockefeller University Press, NY, pp. 962–972.
- Elder, J.F., Jr and Turner, B.J. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol.*, **70**, 297–320.
- Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. and Jasin, M. (1998) Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell. Biol.*, **18**, 93–101.
- Ellis, J. and Bernstein, A. (1989) Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology. *Mol. Cell. Biol.*, **9**, 1621–1627.
- Farve, D. (1992) Improved phenol based method for the isolation of DNA fragments from low melting temperature agarose gels. *BioTechniques*, **13**, 24–26.

- Gangloff, S., Zou, H. and Rothstein, R. (1996) Gene conversion plays the major role in controlling the stability of large tandem repeats in yeast. *EMBO J.*, **15**, 1715–1725.
- Harris, R.S., Kong, Q. and Maizels, N. (1999) Somatic hypermutation and the three R's: repair, replication and recombination. *Mutat. Res.*, **436**, 157–178.
- Hillis, D.M., Moritz, C., Porter, C.A. and Baker, R.J. (1991) Evidence for biased gene conversion in concerted evolution of ribosomal DNA. *Science*, **251**, 308–310.
- Hipeau-Jacquotte, R., Brutlag, D.L. and Bregegere, F. (1989) Conversion and reciprocal exchange between tandem repeats in *Drosophila melanogaster*. *Mol. Gen. Genet.*, **220**, 140–146.
- Holbeck, S.L. and Strathern, J.N. (1997) A role for REV3 in mutagenesis during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics*, **147**, 1017–1024.
- Iype, L.E., Inman, R.B. and Cox, M.M. (1995) Blocked RecA protein-mediated DNA strand exchange reactions are reversed by the RuvA and RuvB proteins. *J. Biol. Chem.*, **270**, 19473–19480.
- Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, **6**, 136–145.
- Jeffreys, A.J., Murray, J. and Neumann, R. (1998) High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell*, **2**, 267–273.
- Langin, T., Hamza, H., Haedens, V. and Rossignol, J.L. (1988a) Reciprocal exchanges instigated by large heterologies in the b2 gene of *Ascomobolus* are not associated with long adjacent hybrid DNA stretches. *Genetics*, **119**, 329–336.
- Langin, T., Haedens, V. and Rossignol, J.L. (1988b) Hybrid DNA extension and reciprocal exchanges: alternative issues of an early intermediate during meiotic recombination? *Genetics*, **119**, 337–344.
- Liao, D. and Weiner, A.M. (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the *RNU2* locus) does not prevent rapid diversification of the (CT)_n-(GA)_n microsatellite embedded within the U2 repeat unit. *Genomics*, **30**, 583–593.
- Liao, D., Pavelitz, T., Kidd, J., Kidd, K.K. and Weiner, A.M. (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J.*, **16**, 588–598.
- Liao, D., Pavelitz, T. and Weiner, A.M. (1998) Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. *J. Mol. Evol.*, **46**, 649–660.
- Matera, A.G., Weiner, A.M. and Schmid, C. (1990) Structure and evolution of the U2 snRNA multigene family in primates, gene amplification under natural selection. *Mol. Cell. Biol.*, **10**, 5876–5882.
- May, C.A., Jeffreys, A.J. and Armour, J.A. (1996) Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (*D16S309*). *Hum. Mol. Genet.*, **5**, 1823–1833.
- Mayer, J., Meese, E. and Mueller-Lantzsch, N. (1998) Human endogenous retrovirus K homologous sequences and their coding capacity in Old World primates. *J. Virol.*, **72**, 1870–1875.
- Murti, J.A., Bumbulis, M. and Schimenti, J.C. (1992) High-frequency germ line gene conversion in transgenic mice. *Mol. Cell. Biol.*, **12**, 2545–2552.
- Nag, D.K. and Petes, T.D. (1990) Meiotic recombination between dispersed repeated genes is associated with heteroduplex formation. *Mol. Cell. Biol.*, **10**, 4420–4423.
- Nagylaki, T. (1984) Evolution of multigene families under interchromosomal gene conversion. *Proc. Natl Acad. Sci. USA*, **81**, 3796–3800.
- Nagylaki, T. and Petes, T.D. (1982) Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics*, **100**, 315–337.
- Neuhausen, S.L. et al. (1994) A P1-based physical map of the region from *D17S776* to *D17S78* containing the breast cancer susceptibility gene *BRCA1*. *Hum. Mol. Genet.*, **3**, 1919–1926.
- Ohta, T. (1976) Simple model for treating evolution of multigene families. *Nature*, **262**, 74–76.
- Ohta, T. and Dover, G.A. (1983) Population genetics of multigene families that are dispersed into two or more chromosomes. *Proc. Natl Acad. Sci. USA*, **80**, 4079–4083.
- Ozenberger, B.A. and Roeder, G.S. (1991) A unique pathway of double-strand break repair operates in tandemly repeated genes. *Mol. Cell. Biol.*, **11**, 1222–1231.
- Parsons, R. et al. (1993) Hypermutability and mismatch repair deficiency in RER⁺ tumor cells. *Cell*, **75**, 1227–1236.
- Pavelitz, T., Rusché, L., Matera, A.G., Scharf, J.M. and Weiner, A.M. (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs *in situ*, without changing the cytological context of the *RNU2* locus. *EMBO J.*, **14**, 169–177.
- Pierce, J.C., Sauer, B. and Sternberg, N. (1992) A positive selection vector for cloning high molecular weight DNA by the bacteriophage P1 system: improved cloning efficacy. *Proc. Natl Acad. Sci. USA*, **89**, 2056–2060.
- Prolla, T.A., Pang, Q., Alani, E., Kolodner, R.D. and Liskay, R.M. (1994) MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. *Science*, **265**, 1091–1093.
- Richard, M., Belmaaza, A., Gusew, N., Wallenburg, J.C. and Chartrand, P. (1994) Integration of a vector containing a repetitive LINE-1 element in the human genome. *Mol. Cell. Biol.*, **14**, 6689–6695.
- Richard, M., Gusew, N., Belmaaza, A. and Chartrand, P. (1997) Homologous junctions formed between a vector and human genomic repetitive LINE-1 elements as a result of one-sided invasion. *Somat. Cell. Mol. Genet.*, **23**, 75–81.
- Rockmill, B., Engebrecht, J.A., Scherthan, H., Loidl, J. and Roeder, G.S. (1995) The yeast *MER2* gene is required for chromosome synapsis and the initiation of meiotic recombination. *Genetics*, **141**, 49–59.
- Schlötterer, C. and Tautz, D. (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol.*, **4**, 777–783.
- Seigneur, M., Bidnenko, V., Ehrlich, S.D. and Michel, B. (1998) RuvAB acts at arrested replication forks. *Cell*, **95**, 419–430.
- Seperack, P., Slatkin, M. and Arnheim, N. (1988) Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. *Genetics*, **119**, 943–949.
- Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science*, **191**, 528–535.
- Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, **365**, 274–276.
- Szostak, J.W. and Wu, R. (1980) Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature*, **284**, 426–430.
- Tracy, S. (1981) Improved rapid methodology for the isolation of nucleic acids from agarose gels. *Prep. Biochem.*, **11**, 251–268.
- Van Arsdell, S.W. and Weiner, A.M. (1984) Pseudogenes for human U2 do not have a fixed site of 3' truncation. *Nucleic Acids Res.*, **12**, 1463–1471.
- Van Valen, L. (1973) A new evolutionary law. *Evol. Theory*, **1**, 1–30.
- Villemure, J. F., Belmaaza, A. and Chartrand, P. (1997) The processing of DNA ends at double-strand breaks during homologous recombination: different roles for the two ends. *Mol. Gen. Genet.*, **256**, 533–538.
- Walsh, J.B. (1987) Persistence of tandem arrays: implications for satellite and simple sequence DNAs. *Genetics*, **115**, 553–567.
- Warburton, P.E. and Willard, H.F. (1995) Interhomologue sequence variation of alpha satellite DNA from human chromosome 17, evidence for concerted evolution along haplotypic lineages. *J. Mol. Evol.*, **41**, 1006–1015.
- Warburton, P.E., Wayne, J.S. and Willard, H.F. (1993) Unequal crossing-over of tandemly repeated DNA: implications for higher-order structural characteristics within centromeric heterochromatin. *Mol. Cell. Biol.*, **13**, 6520–6529.
- Weiner, A.M. and Denison, R.A. (1983) Either gene amplification or gene conversion may maintain the homogeneity of the multigene family encoding human U1 small nuclear RNA. *Cold Spring Harbor Symp. Quant. Biol.*, **47**, 1141–1149.
- Wolff, R.K., Plaetke, R., Jeffreys, A.J. and White, R. (1989) Unequal crossing over between homologous chromosomes is not the major mechanism involved in the generation of new alleles at VNTR loci. *Genomics*, **5**, 382–384.

Received March 16, 1999; revised and accepted May 3, 1999