The EMBO Journal Vol.16 No.3 pp.588–598, 1997

Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the *RNU2* locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion

Daiqing Liao[1,2], Thomas Pavelitz[1], Judith R.Kidd[3], Kenneth K.Kidd[3] and Alan M.Weiner[1,3]

Departments of [1]Molecular Biophysics and Biochemistry and [3]Genetics, Yale University School of Medicine, New Haven, CT 06510-8024, USA

[2]Corresponding author

We have surveyed the tandemly repeated genes encoding U2 snRNA in a diverse panel of humans. We found only two polymorphisms within the U2 repeat unit: a *Sac*I polymorphism (alleles *SacI*$^+$ or *Sac*I$^-$) and a CT microsatellite polymorphism (alleles CT$^+$ or CT$^-$). Surprisingly, individual U2 tandem arrays are entirely *Sac*I$^+$ or *Sac*I$^-$, and entirely CT$^+$ or CT$^-$, although the *Sac*I and CT alleles can occur in any combination. We also found that polymorphisms in the left and right junction regions flanking the tandem array fall into only two haplotypes (JL$^+$ and JL$^-$, JR$^+$ and JR$^-$). Most surprisingly, JL$^+$ is always associated with JR$^+$, and JL$^-$ with JR$^-$. Thus individual U2 arrays do not exchange flanking markers, despite independent assortment and subsequent homogenization of the *Sac*I and CT alleles within the U2 repeat units. We propose that the primary driving force for concerted evolution of the tandem U2 genes is intrachromosomal homogenization; interchromosomal genetic exchanges are much rarer, and reciprocal nonsister chromatid exchange apparently does not occur. Thus concerted evolution of the U2 tandem array occurs *in situ* along a chromosome lineage, and linkage disequilibrium between sequences flanking the U2 array may persist for long periods of time.

Keywords: concerted evolution/human genetic diversity/linkage disequilibrium/multigene family/recombination

Introduction

Tandemly repeated multigene families constitute a significant fraction of most metazoan genomes. For example, the multigene families encoding the large rRNAs, 5S rRNA and the abundant U1 and U2 small nuclear RNAs (snRNAs) together account for ~2% of the human genome. The tandemly repeated multigene families encoding rRNA (Arnheim *et al.*, 1980) and U2 snRNA (Pavelitz *et al.*, 1995) are known to undergo concerted evolution in humans and primates, i.e. individual repeat units of a tandem array are very similar (if not identical) within each species, but differ significantly from the orthologous repeat units of closely related species (for a review, see Elder and Turner, 1995). Homogenization of a tandem array could, in principle, occur by cycles of unequal crossover (Smith, 1976), gene conversion (Dover, 1982) or contraction and expansion of the array (Ozenberger and Roeder, 1991). In lower eukaryotes, the data appear to be consistent with aspects of each model: the yeast ribosomal DNA (rDNA) locus undergoes frequent mitotic and meiotic sister chromatid exchange (Petes, 1980; Szostak and Wu, 1980) as well as gene conversion (Rockmill *et al.*, 1995; Gangloff *et al.*, 1996). Although highly informative studies of rDNA arrays have been reported in flies, mice and humans (Seperack *et al.*, 1988; Schlötterer and Tautz, 1994; reviewed by Elder and Turner, 1995), the mechanisms of concerted evolution in metazoans have been largely inferred from theoretical studies (Ohta, 1976; Smith, 1976; Dover, 1982; Nagylaki and Petes, 1982; Ohta and Dover, 1983; Nagylaki, 1984; Walsh, 1987) because the experimental analysis of tandemly repeated genes has proved so challenging. In particular, the various mechanisms proposed to account for concerted evolution could not be distinguished clearly in the absence of detailed information correlating genetic changes within a tandem array with changes in both flanking sequences.

To understand the molecular mechanism(s) of concerted evolution in higher eukaryotes, we have undertaken a detailed genetic analysis of the tandemly repeated U2 snRNA genes (the *RNU2* locus) in human populations. The relatively small size and uniform structure of the *RNU2* locus provide an excellent opportunity to investigate the mechanisms of concerted evolution. The human *RNU2* locus maps to a single chromosomal site at 17q21–q22 (Hammarstrøm *et al.*, 1984; Lindgren *et al.*, 1985), and the number of 6.1 kb repeat units per U2 tandem array varies from six to >30 (Pavelitz *et al.*, 1995). Thus intact *RNU2* arrays range in size from ~37 to >200 kbp, and the two intact *RNU2* arrays from a diploid genome can almost always be resolved and physically purified by field inversion gel electrophoresis (FIGE; Pavelitz *et al.*, 1995). Within a single tandem array, each of the repeat units is apparently identical except for an embedded CT microsatellite, which is slightly heterogeneous because it evolves faster than the U2 repeat unit can be homogenized (Liao and Weiner, 1995).

Other tandemly repeated mammalian genes may not be as well suited for detailed genetic studies as the *RNU2* locus. For example, the human rRNA genes have a larger repeat unit (>43 kb), a longer tandem array (~100 repeats) and the ~500 genes are divided among five non-syntenic arrays (nucleolus organizers) which are highly polymorphic both within and between chromosomes (Seperack *et al.*, 1988; Gonzalez *et al.*, 1988, 1992). Although the tandem repeat unit of the human 5S rRNA genes is quite small (2.3 kb) and the 5S arrays are only slightly polymorphic, detailed genetic studies of the *RN5S* locus would be seriously confounded by the 10-fold excess of closely related, but diverse 5S pseudogenes (Sorensen

© Oxford University Press

and Frederiksen, 1991). In contrast, human U2 genes outnumber the U2 pseudogenes (Dahlberg and Lund, 1988). Tandemly repeated non-coding DNA sequences are also common in eukaryotic genomes, and range from apparently non-functional simple microsatellites to vast tandem arrays with potential centromeric functions (reviewed by Willard, 1990; Charlesworth *et al.*, 1994). These sequences present different problems. For example, although human alphoid satellite DNA evolves concertedly (Warburton and Willard, 1995), the arrays are vast (300 to >5000 kb), have complex internal repeat structures, are present on every chromosome and are polymorphic between chromosomes (Willard, 1990). Similarly, although a great deal has been learned about the concerted evolution of minisatellites with small repeat units (<100 bp) and relatively small array size (Jeffreys *et al.*, 1985, 1994), it is still not clear whether minisatellite arrays provide a good model for larger functional tandem arrays, or require a small repeat unit and/or special sequences.

We have now characterized individual U2 tandem arrays in eight diverse human populations ranging from our African origins to some of the furthest reaches of the human diaspora (Armour *et al.*, 1996; Tishkoff *et al.*, 1996). The analysis depended on our ability to isolate individual U2 tandem arrays from diploid DNA by FIGE, and to recover each individual U2 array or parts thereof by polymerase chain reaction (PCR). This array-specific PCR protocol, in conjunction with genomic blotting, has allowed us to characterize the haplotypes of individual U2 tandem arrays and the chromosomal DNA immediately flanking them. We show that individual U2 arrays are homogeneous for each polymorphic marker examined, although the polymorphic markers within a U2 tandem array can undergo random assortment on an evolutionary time scale. Most remarkably, random assortment and subsequent homogenization of polymorphic markers does not affect or involve flanking chromosomal DNA. Instead, we find that the DNA flanking the U2 tandem array falls into only two haplotypes, and these haplotypes are never disjoined by reciprocal recombination. Our data imply that (i) arraywide gene conversion and/or sister chromatid exchange are the primary mechanisms of concerted evolution in the human *RNU2* locus, (ii) gene conversion (but not reciprocal recombination) is responsible for non-sister chromatid exchange and (iii) non-sister exchange (between homologs) occurs very infrequently if at all compared with intrachromosomal and sister exchange events.

Results

SacI polymorphism of U2 tandem arrays

The sequence of the 6.1 kb U2 repeat unit is quite homogeneous in human populations (Van Arsdell and Weiner, 1984; Matera *et al.*, 1990; Liao and Weiner, 1995; Pavelitz *et al.*, 1995) and DNA polymorphisms within the *RNU2* locus are correspondingly rare. To search for possible restriction fragment length polymorphisms (RFLPs), we digested a panel of diverse human DNAs with >20 different restriction enzymes. Genomic blotting revealed only a single polymorphic *Sac*I site in the U2 repeat unit, and this was due to a transition between A and G at position 4292 (GAACTC in *Sac*I^{-}, GAGCTC in *Sac*I^{+}; see GenBank entry U57614). The *Sac*I polymorphism was found in all populations tested to date, and DNA sequencing confirmed that this polymorphism is due solely to a transition between A and G at position 4292 in all cases examined (see below, and data not shown). These observations strongly suggest that the *Sac*I polymorphism is ancient, and should be informative for tracing recombination and/or gene conversion events during concerted evolution of the human *RNU2* locus.

To study the inheritance of the *Sac*I polymorphism, we examined U2 tandem arrays in an Old Order Amish pedigree that includes 10 members of three generations (Figure 1A). Genomic DNA from Epstein–Barr virus (EBV)-transformed lymphocyte lines derived from each individual was digested by *Sac*I, resolved by agarose gel electrophoresis and probed for the U2 repeat unit (Figure 1B). *Sac*I digestion of a *Sac*I^{+} repeat unit gives rise to three fragments of 2.8, 1.9 and 1.4 kb, whereas a *Sac*I^{-} repeat unit yields two fragments of 4.7 and 1.4 kb (the 1.4 kb fragment does not react with the probe used in Figure 1B). Some individuals proved to be *Sac*I^{+} or *Sac*I^{-} homozygotes, while others were heterozygotes for the *Sac*I polymorphism (Figure 1B). The *Sac*I^{+} or *Sac*I^{-} homozygotes appeared to be pure; we would easily have detected a single *Sac*I^{+} site in an otherwise *Sac*I^{-} array, or a single *Sac*I^{-} repeat in an otherwise *Sac*I^{+} array. The heterozygotes, however, could be explained in either of two ways. Heterozygotes might have two homogeneous U2 tandem arrays, one *Sac*I^{+} and the other *Sac*I^{-}. Alternatively, heterozygotes might have mixed U2 tandem arrays containing both *Sac*I^{+} and *Sac*I^{-} repeat units, perhaps resulting from reciprocal recombination or from patchwise gene conversion between *Sac*I^{+} and *Sac*I^{-} arrays.

To distinguish between these possibilities, we determined the state of the *Sac*I polymorphism in single U2 tandem arrays derived from individuals known by direct genomic blotting to be *Sac*I$^{+/-}$ heterozygotes. The two U2 arrays from each individual were excised from flanking chromosomal DNA by digestion with *Eco*RI (a 'null cutter' which does not cut within the U2 repeat unit), resolved by FIGE and the dried agarose gel ('unblot') was probed with the *Nhe*I–*Nde*I fragment of the U2 repeat in order to locate individual U2 tandem arrays relative to known DNA size markers (see Figure 1C). To determine the state of the *Sac*I polymorphism in each individual U2 array, bands corresponding to individual U2 arrays were excised from the 'unblot' and used as template for array-specific PCR amplification (Liao and Weiner, 1995) of a 721 bp fragment encompassing the polymorphic *Sac*I site. The 721 bp PCR product was then digested with *Sac*I and the products resolved by agarose gel electrophoresis (Figure 2). Note that the PCR primers were >200 bp from the diagnostic *Sac*I site, and the righthand primer falls outside the *Nhe*I–*Nde*I fragment used to probe the 'unblot'; thus *Sac*I^{+} and *Sac*I^{-} fragments will be amplified with equal efficiency, and labeled fragments derived from the *Nhe*I–*Nde*I probe used during 'unblotting' will not be amplified.

We find that individual U2 arrays are either entirely *Sac*I^{+} or entirely *Sac*I^{-} (Figure 2, and also see below). The 721 bp PCR product derived from *Sac*I^{-} U2 arrays was completely resistant to *Sac*I digestion, and almost all of the PCR product from *Sac*I^{+} U2 arrays was cleaved into two fragments of expected length (470 and 251 bp)

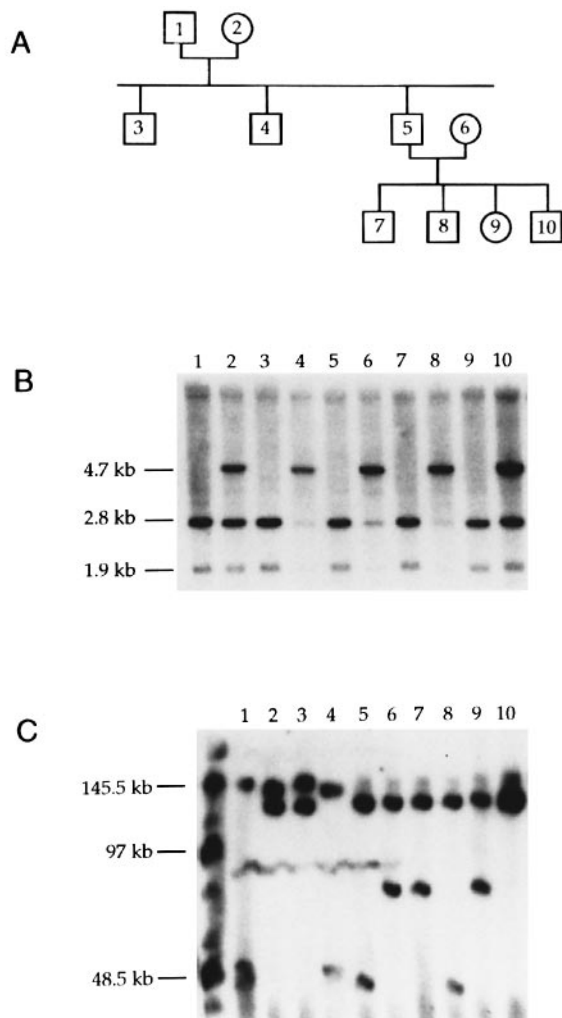

Fig. 1. *Sac*I polymorphism within human U2 tandem arrays. (**A**) Pedigree of 10 members of a large Old Order Amish kindred. (**B**) *Sac*I polymorphism of the U2 tandem arrays in the Old Order Amish pedigree. Genomic DNA from EBV-transformed lymphoblastoid lines (GM5963, GM5961, GM5993, GM5995, GM5927, GM5929, GM5935, GM5937, GM5941 and GM5943 for individuals 1–10, respectively) derived from individuals in (A) was digested with *Sac*I, resolved by conventional agarose gel electrophoresis, and the dried gel ('unblot') probed directly with the *Nhe*I–*Dra*I fragment of the human U2 repeat unit (Pavelitz *et al.*, 1995). The unequal intensity of individual *Sac*I fragments is due to length (and hence copy number) variation between the two U2 arrays. For example, the *Sac*I$^-$ bands are darker than the *Sac*I$^+$ bands in lanes 4 and 8 because the *Sac*I$^-$ array is larger. Copy number variation initially made it difficult to interpret RFLP patterns, because we were unable to distinguish a small *Sac*I$^-$ array from an incompletely digested *Sac*I$^+$ array. Lane numbers from left to right correspond to the numbered individuals in (A). Sizes of the *Sac*I fragments are indicated. (**C**) A genomic unblot of intact U2 tandem arrays in the Old Order Amish pedigree. Intact U2 arrays were released from flanking chromosomal DNA by digestion with *Eco*RI (a 'null cutter' which does not cut within the U2 repeat unit) and the arrays were resolved by field inversion gel electrophoresis (FIGE). Unblotting was carried out as in (B). Markers in the leftmost lane were λ MidRange Marker I (New England Biolabs). Lanes as in (B).

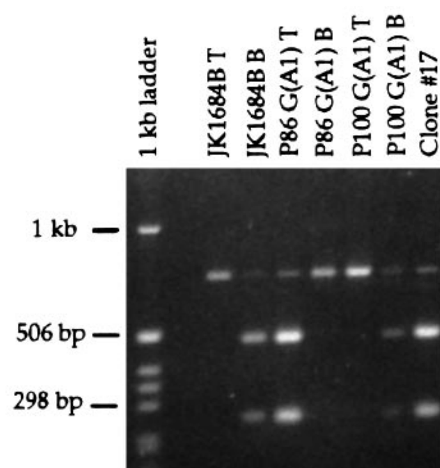

Fig. 2. Individual U2 tandem arrays are homogeneous for the *Sac*I polymorphism. A 721 bp fragment encompassing the *Sac*I$^{+/-}$ site at 4292 within the 6.1 kb U2 repeat unit (from position 3822 to 4543) was amplified from isolated U2 tandem arrays by array-specific PCR (Liao and Weiner, 1995). The PCR products were digested with *Sac*I, and the resulting fragments separated by electrophoresis through a 1.2% agarose gel in the presence of ethidium bromide. Three *Sac*I$^{+/-}$ heterozygotes are analyzed here [JK1684B, P86G(A1) and P100G(A1)]. For each individual, the larger parental array is designated 'T' (top) and the smaller array 'B′ (bottom), e.g. JK1684B T and JK1684B B. Clone #17 is a plasmid containing a *Sac*I$^+$ U2 fragment (Pavelitz *et al.*, 1995). The leftmost lane is a 1 kb ladder (GIBCO-BRL).

(Figure 2). These results are consistent with one of two intriguing scenarios: either (i) there are only two kinds of U2 arrays in modern human populations, and these do not undergo reciprocal recombination with each other, or (ii) individual *Sac*I sites can undergo interconversion between the *Sac*I$^+$ and *Sac*I$^-$ states, followed by rapid homogenization of the entire U2 tandem array.

The small amount of *Sac*I-resistant PCR product derived from the *Sac*I$^+$ arrays [lanes JK1684B B, P86G(A1) T, P100G(A1) B and Clone #17] appears, for three compelling reasons, to be a PCR artefact rather than an indication of U2 arrays containing a mixture of *Sac*I$^+$ and *Sac*I$^-$ repeats. First, although a *Sac*I$^+$ plasmid (Clone #17; Pavelitz *et al.*, 1995) could be digested to completion by *Sac*I, the 721 bp PCR product amplified directly from this plasmid could not (Figure 2, lane Clone #17). Second, a small amount of *Sac*I-resistant 721 bp PCR product was also observed when the template for PCR amplification was genomic DNA from a *Sac*I$^{+/+}$ homozygote whose two U2 arrays were known by direct genomic blotting to be entirely *Sac*I$^+$, i.e. no trace of *Sac*I$^-$ repeats could be detected under conditions where single copy genomic fragments are clearly visible (data not shown). Third, we confirmed that individual U2 arrays are either entirely *Sac*I$^+$ or *Sac*I$^-$ by *Sac*I digestion of individual U2 tandem arrays purified by preparative low melting point agarose gel electrophoresis (data not shown). In addition, we cloned and sequenced several of the amplified PCR products from one individual in each of four diverse populations (Chinese, Mbuti, Melanesian and Surui), confirming in each case that the *Sac*I polymorphism was due to a transition between A and G at position 4292. Occasional nucleotide substitutions were also found in the 721 bp fragment of these diverse ethnic groups. The observed nucleotide substitutions are unlikely to be PCR artefacts because almost every substitution was shared by two sequences from different populations. We can estimate, therefore, that the average sequence divergence among the U2 arrays in these four different human populations

Table I. Haplotypes of U2 tandem arrays and flanking sequences in human populations

DNA source	Array	Array length (kb)	Position in JL			Position in JR		*Sac* I polymorphism	CT microsatellite polymorphism
			–137	–134	+5	+42	+54		
GM5927	top	135	A	T	G	A	A	+	+
	bottom	50	A	T	G	A	A	+	+
GM5929	top	135	C	C	C	T	G	–	n.d.
	bottom	90	A	T	G	A	A	+	n.d.
GM5935	top	135	A	T	G	A	A	+	n.d.
	bottom	90	A	T	G	A	A	+	n.d.
GM5937	top	135	C	C	C	T	G	–	n.d.
	bottom	50	A	T	G	A	A	+	n.d.
DL	top	>200	A	T	G	A	A	+	+
	bottom	145	A	T	G	A	A	+	+
1 (Mbuti)	x	>200	C	C	C	T	G	–	n.d.
2 (Mbuti)	top	>200	A	T	G	A	A	–	+
	bottom	110	A	T	G	A	A	–	+
3 (Mbuti)	top	200	C	C	C	T	A	–	n.d.
	bottom	140	C	C	C	T	A	–	n.d.
JK1684B	top	190	A	T	G	T	A	–	–
	bottom	135	A	T	G	A	A	+	–
P86G(A1)	top	115	A	T	G	A	A	+	+
	bottom	95	A	T	G	A	A	–	+
P86gG(A1)	top	190	A	T	G	–	A	–	–
	bottom	145	A	T	G	A	A	–	+
P100G(A1)	top	85	A	T	G	A	A	–	+
	bottom	40	A	T	G	A	A	+	+
5 (Melanesian)	top	190	C	C	C	T	G	–	–
	bottom	70	C	C	C	T	G	–	–
7 (Surui)	x	>200	C	C	C	T	G	–	n.d.
8 (Surui)	top	190	A	T	G	A	A	+	+
	bottom	170	A	T	G	A	A	+	+

Only polymorphic nucleotides in the left and right junction regions (JL and JR) of the U2 tandem arrays are shown explicitly. The presence or absence of the *Sac*I polymorphism at position 4292 in the U2 tandem array is indicated ('+' or '–'). Informative polymorphisms in the CT microsatellite are also labeled '+' or '–' as described in Figure 3. U2 array sizes were estimated from FIGE-separated *Eco*RI genomic fragments that were visualized by hybridizing U2-specific probe (see Figure 1C). The size standard was MidRange Marker I (New England Biolabs). Cell lines GM5927, GM5929, GM5935 and GM5937 were derived from individuals 5, 6, 7 and 8 of the Amish pedigree shown in Figure 1A. DL is a Chinese individual whose DNA was isolated directly from fresh lymphocytes. Cell lines were from individuals in the Biaka and Mbuti tribes of African pygmy [JK1684B, P86G(A1), P86gG(A1), P100G(A1), 1, 2, 3], a Nasioi (Melanesian, 5) and the Rondonian Surui tribe of South American Indians (7 and 8). When both U2 arrays in an individual were analyzed, the larger and smaller arrays were designated 'top' and 'bottom'; when only one array was analyzed, it is designated 'x'. Nucleotides are numbered according to Pavelitz *et al.* (1995). n.d., not determined.

is <0.3% (based on the sequences of the 721 bp fragment and excluding the hypervariable CT microsatellite).

Length variation of U2 tandem arrays

When excised with the null cutter *Eco*RI, the lengths of intact U2 tandem arrays in the Amish pedigree vary from 50 to 135 kb (Figure 1C). This corresponds to seven to >22 U2 repeat units per array after the size of the junction fragments JL and JR is taken into account (Pavelitz *et al.*, 1995). Such length variation suggests a high level of ongoing recombination within or between *RNU2* loci. To study the length distribution of U2 arrays in a larger sample, we surveyed >80 chromosomes in diverse human populations. We found that the length of individual U2 tandem arrays varies widely from as low as 40 kb (~6 U2 repeats) to ~200 kb (>30 U2 repeats; Table I and data not shown). In the chromosomes surveyed, 57% of the U2 arrays were between 100 and 200 kb, 32% were 40–100 kb and 11% were longer than the 200 kb resolution limit of our FIGE regime. While we never detected U2 arrays smaller than 40 kb, extremely long U2 arrays (>250 kb) are also very rare (data not shown). These observations indicate that U2 tandem arrays undergo frequent recombination among themselves to generate high levels of length polymorphism within each population. The remarkable length polymorphism of the U2 arrays also underscores the importance of a gene dosage compensation mechanism(s) that can maintain a relatively fixed level of U2 snRNA over a 4-fold or greater range of U2 gene dosage (A.D.Bailey and A.M.Weiner, unpublished; see also Mangin *et al.*, 1985).

Knowing that each U2 array is homogeneous for the *Sac*I polymorphism (Figures 1 and 2), we next examined the inheritance of U2 tandem arrays in the same Amish family to determine whether U2 array lengths were stable between generations (Figure 1C, lane 4). The two U2 arrays in each individual (Figure 1A) were resolved by FIGE (Figure 1C), located by probing an 'unblot' with the *Nhe*I–*Nde*I fragment, and the state of the polymorphic *Sac*I site in the individual U2 arrays assayed by array-specific PCR, as in Figure 2, or by Southern blotting (see above). We found that inheritance of both the number of U2 repeats per array and the *Sac*I state of each array is strictly Mendelian. For example, individual #8 (Figure 1C, lane 8) inherited the upper *Sac*I^{-} array (135 kb) from the mother (#6) and the lower *Sac*I^{+} array (50 kb) from the father (#5), resulting in a *Sac*I$^{+/-}$ heterozygote (Figure 1B, lane 8). Thus, it is likely that the frequency of

```
              1                                                50
GM5927-CT-10  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CCTTCC
     DL-CT-5  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC
    DL-CT-12  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC
    H6-CT-15  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT TT..CTTTCC
    H6-CT-23  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC
    Mb-#2-24  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC
    Mb-#2-22  CTCTCTCTCA CTCTCTCTCT CCTCTCTCCT GTCTCTCTCT CT..CTTTCC
    Mb-#2-20  CTCTCTCTCA ..CTCTCTCT CCTCTCTCCT GTCTCTCTCT CT..CTTTCC
    Me-#5-26  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC
    Me-#5-25  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CTCTCTTTCC
     WJ-CT-1  CTCTCTCTCT ........CT CCTCTCTC.T GTCTCTCTCT CT..CTTTCC

              51                                              100
GM5927-CT-10  C..TCTCTCT CTCTCTTCCC CTCCC.CCGC CTCTCCCTCG C....TCTTT
     DL-CT-5  C..TCTCTCT CTCTCTTCCC TCCCCCCCGC CTCTCCCTCG C....TCTTT
    DL-CT-12  C..TCTCTCT CTCTCTTCCC CCCCTCCCGC CTCTCCCTCG C....TCTTT
    H6-CT-15  C..TCTCTCT CTCTTCCCCC CCCCC..CGC CTCTCCCTCG C....TCTTT
    H6-CT-23  CTCTCTCTCT CTCTCTTCCC CCCCC...GC CTCTCCCTCG C....TCTTT
    Mb-#2-24  C..TCTCTCT CTCTCTTCCA CCCCC.CCGC CTCTCCCTCG C... TCTTT
    Mb-#2-22  C..TCTCTCT CTCTCTTTCC CCCCCCCCGC CTCTCCCTCG C....TCTTT
    Mb-#2-20  C..TCTCTCC CTCTCTTCCA CCCCC.CCGC CTCTCCCTCG C....TCTTT
    Me-#5-26  C......... .......CCC CCCCC.CCGC CTCTCCCTCG CTCTCTCTTT
    Me-#5-25  CC........ .......CCC CCCCC.CCGC CTCTCCCTCG CTCTCTCTTT
     WJ-CT-1  C......... .......CCC CCCTCCCCGC CTCTCCCTCG CTCTCTCTTT

              101                            133
GM5927-CT-10  AGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
     DL-CT-5  AGGTTTCCCC TACCCCCTCC CAAGTTCTGG GGT
    DL-CT-12  AGGTTTCCCC CACTCCCTCC CAAGTTCTGG GGT
    H6-CT-15  AGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
    H6-CT-23  AGGTTTCCCT CACCCCCTCC CAAGTTCTGG GGT
    Mb-#2-24  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
    Mb-#2-22  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
    Mb-#2-20  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
    Me-#5-26  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
    Me-#5-25  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
     WJ-CT-1  TGGTTTCCCC CACCCCCTCC CAAGTTCTGG GGT
```

Fig. 3. CT microsatellites amplified from individual human U2 tandem arrays by array-specific PCR. Microsatellites were amplified, cloned and sequenced as described (Liao and Weiner, 1995). Representative CT microsatellites from individual U2 arrays (GM5927-CT-10; DL-CT-5, 12; H6-CT-15, 23, Mb-#2-CT-20, 22, 24; Me-#5-25, 26 and WJ-CT-1) were aligned. DNA sequences of the CT microsatellites (GM5927-CT-10; DL-CT-5, 12; H6-CT-15, 23) as well as WJ-CT-1 are published sequences (Liao and Weiner, 1995; Pavelitz *et al.*, 1995). All CT microsatellites are from a single chromosome. Cell lines GM5927 was as described in Table I. Mb and Me corresponded to cell lines 3 (Mbuti) and 5 (Melanesian) in Table I, respectively. Five or more clones were sequenced in each case, and the CT microsatellites from individual U2 arrays consistently displayed a CT^+ or CT^- allele as described in the text. As shown previously (Liao and Weiner, 1995), the observed CT polymorphism cannot be due to a PCR artifact.

recombination among U2 arrays is modest compared with some hypervariable human minisatellite loci (Jeffreys *et al.*, 1994).

CT microsatellite polymorphism

A large $(CT)_n{\cdot}(GA)_n$ dinucleotide repeat (the CT microsatellite, where n $\approx$70) lies downstream of the U2 snRNA coding region in each 6.1 kb U2 repeat unit (Liao and Weiner, 1995; Pavelitz *et al.*, 1995). Unlike most of the U2 repeat unit (which is homogeneous) and the *Sac*I polymorphism (which appears to be dimorphic), the CT microsatellite is highly polymorphic in length and sequence, both within individual U2 tandem arrays and within populations (Liao and Weiner, 1995). We wondered, therefore, if the CT microsatellite polymorphism could serve as an informative marker for studying recombination between individual U2 tandem arrays. Using array-specific PCR (Liao and Weiner, 1995), we cloned, sequenced and typed the CT microsatellites within various individual U2 tandem arrays. Surprisingly, although the CT microsatellite exhibits both length and sequence polymorphism within individual U2 tandem arrays, two regions of the CT microsatellite were found to vary in an array-specific fashion, i.e. all repeats within an individual U2 array share the same CT allele (Figure 3). We term one of these array-specific alleles CT^- (a deletion of 14–15 nucleotides between positions 52 and 67) and the other array-specific alleles CT^+ (a deletion of four nucleotides, or two CT repeats, between positions 92 and 95). [In the numbering system used in Figure 3 of Liao and Weiner (1995), these two deletions are located between positions 183 and 198, and positions 223 and 226, respectively.] Thus, in addition to the *Sac*I polymorphism, these two CT microsatellite alleles are also informative markers for tracing recombination events within the *RNU2* locus.

The SacI and CT polymorphic markers assort independently

As described above (Figures 2 and 3), individual U2 arrays are always homogeneous for the *Sac*I dimorphism (*Sac*I$^+$ or *Sac*I$^-$) and the CT microsatellite polymorphism (CT^+ or CT^-). Intriguingly, comparison of individual U2 arrays indicates that the *Sac*I and CT microsatellite polymorphisms exhibit strong disequilibrium in non-African populations, but no disequilibrium among African populations. In all typed *RNU2* loci from non-African populations, the *Sac*I$^+$ polymorphism is associated with the CT^+ polymorphism, and the *Sac*I$^-$ polymorphism with the CT^- polymorphism (see Table I), but either *Sac*I allele can be found in combination with either CT allele in the

RNU2 loci of diverse African populations. In every case, however, the reassorted alleles are homogeneous throughout the entire U2 tandem array, for example every repeat in a *Sac*I$^+$,CT$^-$ array is *Sac*I$^+$ and CT$^-$. Admittedly, 'independent assortment' has been defined classically as assortment of alleles following a single meiosis, but the term also accurately and conveniently describes the assortment of the *Sac*I and CT alleles on an evolutionary time scale.

Junction haplotypes associated with individual U2 arrays

Independent assortment of the *Sac*I and CT microsatellite polymorphisms could occur by repeated cycles of reciprocal recombination, or by gene conversion or by a combination of these two mechanisms. We therefore sought polymorphic markers in regions immediately flanking individual U2 tandem arrays so that we could distinguish reciprocal recombination events (which would lead to the exchange of flanking markers) from gene conversion (which could, in principle, leave the flanking markers untouched).

We amplified, cloned, sequenced and typed the 'left' and 'right' junction fragments (Pavelitz *et al.*, 1995; JL and JR) of many individual U2 tandem arrays by array-specific PCR (Liao and Weiner, 1995). [We now know that JL is centromeric based on analysis of ordered P1 genomic clones (D.Liao and A.M.Weiner, unpublished) developed for mapping the *BRCA1* locus (Neuhausen *et al.*, 1994).] Informative polymorphic sites were found in both junction regions (Table I). These include –137 (A/C), –134 (T/C) and +5 (G/C) for the left junction JL, as well as +42 (A/T) and +54 (A/G) for the right junction JR. (The nucleotide coordinates are according to Pavelitz *et al.*, 1995; positions –137 and –134 of the left junction and positions +42 and +54 of the right junction lie outside of the U2 tandem array.) Most interestingly, sequence variants in both junction regions fell into only two haplotypes; the A(–137)T(–134)G(+5) haplotype in JL was always associated with the A(+42)A(+54) haplotype in JR, whereas the C(–137)C(–134)C(+5) haplotype in JL was always associated with the T(+42)G(+54) haplotype in JR. Sequences flanking the U2 tandem array are therefore in complete linkage disequilibrium; no evidence can be seen for exchange of flanking DNA during interchromosomal recombination between U2 tandem arrays. We conclude that gene conversion, rather than reciprocal crossover, is likely to be responsible for interchromosomal recombination within the human *RNU2* locus.

Types of human U2 tandem arrays

The data in Table I allow us to classify human *RNU2* loci into five major types (also shown schematically in Figure 4). Type I is a *Sac*I$^-$/CT$^-$ U2 tandem array, flanked by the C(–137)C(–134)C(+5) and T(+42)G(+54) haplotypes in the left and right junction regions, respectively. Types II–V all share the same left and right junction haplotypes, namely A(–137)T(–134)G(+5) and A(+42)A(+54), but the haplotype of each U2 tandem array is distinct; these are *Sac*I$^+$/CT$^+$ for type II, *Sac*I$^+$/CT$^-$ for type III, *Sac*I$^-$/CT$^+$ for type IV and *Sac*I$^-$/CT$^-$ for type V. Two minor variations in the right junction regions were found. A(+42)

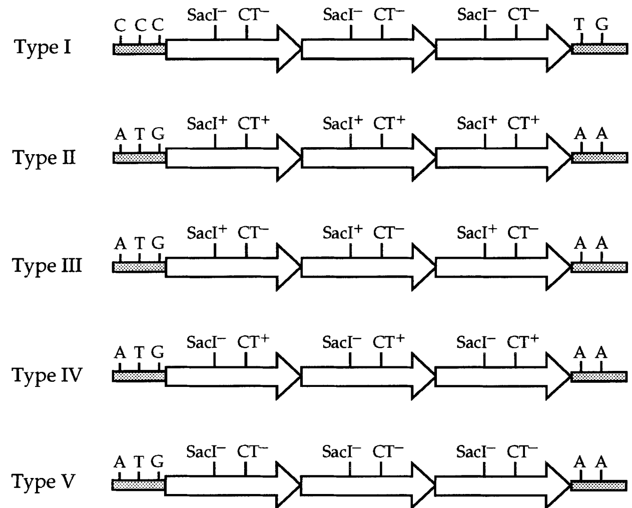

Fig. 4. Schematic representation of the different types of U2 tandem arrays found in human populations. The 6.1 kb U2 repeat unit is shown as a hollow arrow, and flanking sequences as shaded rectangles. *Sac*I$^{+/-}$ and CT$^{+/-}$ are polymorphic markers within the U2 tandem array. Flanking polymorphic sites are indicated by specific nucleotides at each position. For simplicity, the U2 tandem array is shown as containing an integral number of U2 repeat units, as is nearly the case (Pavelitz *et al.*, 1995); flanking sequences are not drawn to scale. Type I and II arrays are the predominant *RNU2* haplotypes in non-African populations, whereas all *RNU2* haplotypes were found in the African populations. Length variations in each type of U2 array were observed (Table I, and data not shown).

was found to be deleted in the right junction of a U2 array in P86gG(A1), and the haplotype of the right junction is T(+42)A(+54) in the two U2 arrays in #3 (Mbuti) and one array in JK1684B. The simplest explanation for this haplotype is that it resulted from a reciprocal crossover event between the A(+42)A(+54) and T(+42)G(+54) haplotypes within the sequences immediately flanking a U2 tandem array. Surprisingly, we have found thus far only type I and type II *RNU2* loci in non-African populations from diverse geographic locations, but all were represented in African populations.

Discussion

There is considerable debate about the mechanisms that are responsible for concerted evolution of tandemly repeated multigene families. Several DNA turnover mechanisms such as unequal crossing over (Smith, 1976), gene conversion (Dover, 1982; Weiner and Denison, 1983; Hillis *et al.*, 1991) and even transposon-mediated gene conversion (Thompson-Stewart *et al.*, 1994) may participate in concerted evolution. These mechanisms may operate differently for different multigene families or even for the same family in different species. Current discussions of concerted evolution suggest that it is essentially a stochastic process, with homogenization of particular haplotypes within a population being achieved by continuous exchange of repeats over a considerable period of time. Thus one would expect that variant repeats, corresponding to intermediates in an ongoing process, would be distributed throughout the multigene family, either randomly or in groups reflecting units of genetic exchange (Dover *et al.*, 1993). However, these discussions often remain

speculative and a more detailed understanding of the mechanism for concerted evolution is clearly needed.

To understand the molecular mechanism of concerted evolution, we have undertaken detailed genetic analyses of the tandemly repeated U2 snRNA genes in diverse human populations as well as in various non-human primates (Liao and Weiner, 1995; Pavelitz *et al.*, 1995). We show here that (i) polymorphic markers can be found among human *RNU2* loci, but these markers are homogeneous within all repeat units of each individual U2 tandem array, despite variation from five to >30 copies in the number of U2 repeat units per array; (ii) U2 tandem arrays exhibit only two common combinations of flanking haplotypes, and no reciprocal exchange between these two tightly associated haplotypes was found thus far; and (iii) polymorphic markers within U2 repeat unit tandem arrays appear to assort independently on an evolutionary time scale without affecting the tight association of flanking haplotypes. We discuss the implications of these findings for the mechanisms of concerted evolution below.

Concerted evolution of the RNU2 locus is driven primarily by intrachromosomal recombination

Concerted evolution of tandemly repeated multigene families must involve two distinct processes: intrachromosomal and interchromosomal exchange. Intrachromosomal recombination homogenizes individual tandem arrays within a single chromosomal lineage, whereas interchromosomal recombination is required to homogenize all tandem arrays within the population. The relative frequency of these two processes could, in principle, be determined by comparing the overall level of homogeneity observed in individual tandem arrays, and in tandem arrays derived from the population as a whole, if genetic drift is insignificant. Our data indicate that the *Sac*I and CT microsatellite polymorphisms are always homogeneous within all repeat units of an individual U2 tandem array, but that polymorphisms between U2 arrays are easily detected (Figure 4). Thus intrachromosomal (within-array) homogenization is far more frequent than interchromosomal (between-array) homogenization. The homogeneity of individual U2 arrays further suggests that intrachromosomal sequence homogenization is not only rapid, but proceeds to completion. Indeed, with the sole (and ironic) exception of the aberrant U2 repeat sequenced by Pavelitz *et al.* (1995), we have never detected a variant repeat within a U2 tandem array. Thus intrachromosomal homogenization must be considerably more rapid than the mutation rate, or heterogeneity would accumulate throughout the repeat unit as it clearly does in the CT microsatellite (Liao and Weiner, 1995). It is even possible that intrachromosomal homogenization could be achieved quite rapidly, perhaps within one or a few meioses or mitoses, although the actual number cannot be determined in the absence of quantitative data for the mutation and recombination rates.

One could argue that the homogeneity of individual U2 arrays we have observed might simply reflect a sampling error, because there may be only a limited number of U2 haplotypes in non-African populations, and these haplotypes may have not diverged sufficiently to generate detectable sequence heterogeneity within individual U2 tandem arrays. However, this scenario is unlikely because (i) individual U2 tandem arrays are also homogenous in African populations, and (ii) length variations observed in all five major *RNU2* types in African populations, and in both type I and II *RNU2* loci in non-African populations, indicate that all these *RNU2* loci continue to undergo genetic exchange (Table I). Furthermore, we have found previously that the orthologous U2 tandem array in baboon consists of 11 kb repeats, whereas the U2 tandem arrays in human, chimpanzee, gorilla, orangutan and gibbon consist of 6 kb repeats; the 5 kb difference represents deletion of a provirus from the ancestral 11 kb repeat unit, leaving behind a solo long terminal repeat (LTR) in all the orthologous 6 kb repeat units that descended from it (Pavelitz *et al.*, 1995). Thus concerted evolution of the primate *RNU2* locus can effectively homogenize both insertions and deletions as large as 5 kb. Taken together, these observations suggest that U2 tandem arrays are dynamic and undergo continuous sequence homogenization.

Our conclusion that intrachromosomal genetic exchange is the primary driving force for concerted evolution of the *RNU2* locus is consistent with a growing body of evidence in other systems. For example, different rRNA arrays in interbreeding populations of *Drosophila melanogaster* are homogenized for different variants (Schlötterer and Tautz, 1994), and linkage disequilibrium among variants of the rDNA loci in humans has also been observed (Seperack *et al.*, 1988). Furthermore, the presence of extensive haplotype-specific sequence variations in tandemly repeated human alphoid satellite DNA suggests that concerted evolution of alphoid satellites also occurs along haplotypic lineages (Warburton and Willard, 1995). Although intraallelic as well as interallelic recombination events are involved in rapid evolution of human minisatellite loci (Buard and Vergnaud, 1994; Jeffreys *et al.*, 1994), the relative homogeneity of these minisatellites may reflect recent expansion rather than (or perhaps in addition to) active homogenization. Thus, intrachromosomal genetic exchanges appear to be the primary driving force for concerted evolution in different tandemly repeated multigene families.

An especially intriguing possibility is that high rates of intrachromosomal recombination may reflect emerging connections between recombination and DNA repair. Specifically, sister chromatids are preferred over homologs as substrates for mitotic recombinational repair in yeast (Kadyk and Hartwell, 1992, 1993), perhaps suggesting that repair of DNA damage could be a major mechanism driving concerted evolution in metazoan systems. Low rates of interchromosomal recombination compared with intrachromosomal recombination have also been observed in mouse somatic cells (Shulman *et al.*, 1995). Low rates of interchromosomal recombination might correlate with the cytological observation that homologs usually reside in different regions in the prometaphase nucleus (Nagele *et al.*, 1995). In this context, it is important to recognize that concerted evolution may reflect a combination of meiotic and mitotic events. Although meiotic events are commonly thought of as the source of all heritable genetic variation in humans, any of the many mitotic events that occur during expansion of germline precursors could also contribute to concerted evolution. Indeed, although both

inter- and intrachromosomal recombination occur at high frequency in the mouse germline, intrachromosomal gene conversion is ~10 times more frequent than interchromosomal events (Murti *et al.*, 1992, 1994), consistent with our data suggesting that intrachromosomal recombination plays the major role in concerted evolution of tandemly repeated genes.

In principle, either intrachromatid gene conversion or unequal sister chromatid exchange (USCE) could account for intrachromosomal recombination during concerted evolution. The homogeneity of human U2 tandem arrays prevents us from distinguishing the relative contributions of these two mechanisms to concerted evolution of the *RNU2* locus, but this can be done experimentally for tandem arrays in yeast using appropriately marked sequences (Jinks-Robertson and Petes, 1993). USCE is certainly the simplest explanation for the observed variation in copy number from five to >30 U2 repeat units per array, but intrachromatid mechanisms cannot be rigorously excluded. Indeed, intrachromatid conversion is often associated with crossovers in yeast (Jinks-Robertson and Petes, 1993), suggesting that intrachromatid homogenization events could also contribute to the observed length variation of human U2 tandem arrays. Alternatively, increases and decreases in array size might reflect polymerase slippage or unequal exchange between replicating sister strands as proposed by Lovett *et al.* (1993), although slippage may be more prevalent during replication of simple sequence repeats such as microsatellites (Schlötterer and Tautz, 1992).

Gene conversion is responsible for interchromosomal recombination

Although less frequent than intrachromosomal homogenization events, interchromosomal recombination must occur sufficiently often to explain why the tandem repeat units of the U2 (Matera *et al.*, 1990) and rDNA arrays (Arnheim *et al.*, 1980) are more similar within each species than between species. In fact, genetic exchange between rDNA arrays on non-homologous chromosomes has been documented in primates (Arnheim *et al.*, 1980) as well as in *Drosophila* (Coen and Dover, 1983) and, more recently, interallelic exchange of blocks of repeats has also been observed in some human minisatellite arrays (Jeffreys *et al.*, 1994). The most likely mechanisms for interchromosomal recombination are reciprocal crossover and/or gene conversion, and these two mechanisms can be distinguished easily if flanking polymorphic markers are known. We therefore identified a number of informative polymorphic markers in regions immediately flanking the U2 tandem array, and then used these flanking markers to test for reciprocal recombination between arrays located on homologous (non-sister) chromatids. Surprisingly, only two kinds of U2 flanking haplotypes were found, and these were in complete disequilibrium despite near equilibrium of polymorphic markers within U2 tandem arrays themselves (i.e. the *Sac*I and CT alleles can be found in any combination; Figures 4 and 5). Thus, genetic exchange within a U2 tandem array does not appear to involve exchange of flanking polymorphic markers, and this argues strongly that interchromosomal recombination is accomplished by gene conversion without reciprocal exchange. These conclusions are fully consistent with the growing body of data on physical and linkage maps of the interval spanning the *RNU2* locus which indicate that the region does not contain a recombination hotspot (e.g. Dib *et al.*, 1996). Our conclusions resemble those of Hillis *et al.* (1991) who demonstrated that a homogeneous rDNA tandem array of one haplotype was replaced consistently by another homogeneous haplotype in the lizard *Heteronotia binoei*. No mosaic or recombinant rDNA arrays containing mixed haplotypes were observed, leading Hillis *et al.* (1991) to conclude that rapid, biased gene conversion, rather than reciprocal recombination, must be responsible for concerted evolution of these rDNA arrays. Similarly, exchange of flanking markers does not accompany the high levels of recombination observed in human hypervariable minisatellites (Wolff *et al.*, 1989; Jeffreys *et al.*, 1994). Interallelic gene conversion may, therefore, be a general mechanism for interchromosomal recombination between tandemly repeated sequences.

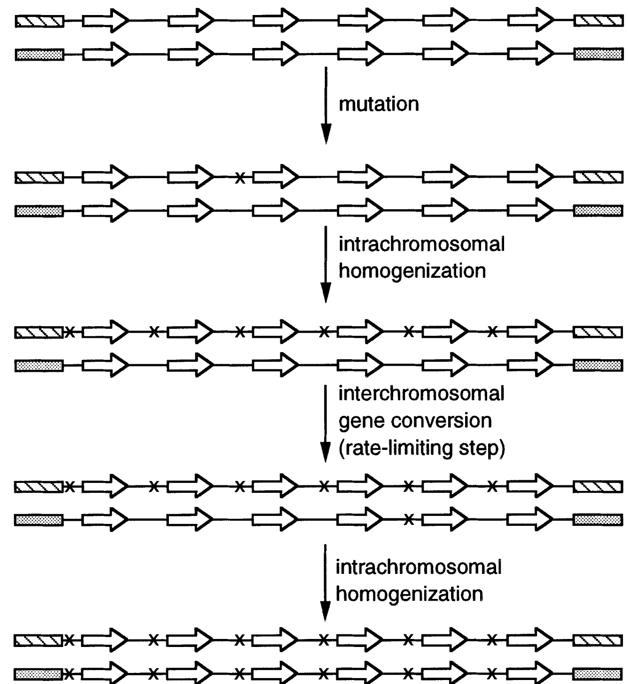

Fig. 5. A model for concerted evolution of the *RNU2* locus in humans and primates. The tandemly repeated U2 arrays on two homologous chromosomes are depicted together with the flanking chromosomal DNA sequences. U2 snRNA coding regions are shown as hollow arrows, spacer sequences as lines and flanking chromosomal DNA as rectangles (cross-hatched and shaded). The chromosomal flanks of the two tandem arrays are labeled differently (cross-hatched or shaded) to emphasize that these flanks exhibit two distinct, tightly associated haplotypes (Table I). One repeat unit in a particular array then acquires a mutation ('X'). The mutation is fixed rapidly within this original array by intrachromosomal homogenization mechanisms, presumably including intrachromatid and unequal sister chromatid recombination. The mutation is then spread to the homologous non-sister array by interallelic genetic exchange, and finally the mutation is fixed throughout the second array by additional rounds of intrachromosomal homogenization. Intrachromosomal homogenization must be much more frequent than interallelic genetic exchange, because individual U2 tandem arrays were homogeneous for all polymorphic markers. Gene conversion is more likely to be responsible for interchromosomal exchange than unequal crossing over, because no exchange of flanking polymorphic markers (cross-hatched and shaded) was observed despite the fact that ongoing genetic activity at the *RNU2* locus is sufficient to generate significant length variation in both African and non-African populations.

The RNU2 gene structure and the mechanism of concerted evolution

Concerted evolution of the primate *RNU2* locus has occurred *in situ* over the past 35 million years, i.e. without apparent cytological movement of the locus, and this suggests that concerted evolution may be facilitated by *cis*-acting elements located within the locus itself, rather than in the flanks (Pavelitz *et al.*, 1995). Potential *cis*-acting sequence elements identified within the U2 repeat include a solo LTR (Pavelitz *et al.*, 1995; D.Liao, T.Pavelitz and A.M.Weiner, submitted), the CT microsatellite (Liao and Weiner, 1995) and the U2 transcription unit itself (Bailey *et al.*, 1995). We (Liao and Weiner, 1995) and others (Htun *et al.*, 1985) have suggested that the CT microsatellite may provide a DNA structure (a 'zipper' sequence) for initiating repeated rounds of recombination and/or gene conversion. Interestingly, a GT microsatellite is found in the 2.2 kb repeat unit of human 5S rRNA arrays (Sorensen and Frederiksen, 1991), and a complex CT-like microsatellite is found in the 43 kb repeat unit of human rDNA arrays (GenBank accession No. U13369). Simple sequence repeats have been proposed to play a similar role in the concerted evolution of protein-coding multigene families in silk moth *Bombyx mori* (Hibner *et al.*, 1991). Alternatively, the CT microsatellite may stimulate recombination by serving as a 'magnet' for repair enzymes instead of a 'zipper' for initiating recombination. Dinucleotide repeats are difficult to replicate accurately, and the resulting replication slippage errors are substrates for the mismatch repair machinery (Parsons *et al.*, 1993). Just as a stalled transcription complex can trigger efficient repair on the template strand (transcription-coupled repair, Mellon *et al.*, 1996), so a stalled replication complex may trigger 'replication-coupled repair' by attracting repair enzymes which in turn stimulate recombination. Such replication-coupled DNA repair mechanisms could cause a pair of replicating tandem arrays to align out of register, and subsequent resolution of the misaligned structure could then lead to contraction or expansion of a tandem array (Lovett *et al.*, 1993).

Remarkably, hotspots of meiotic recombination in the mouse major histocompatibility complex (MHC) also contain a CT-like microsatellite DNA as well as sequences similar to the LTR of one type of murine retrotransposon (Shiroishi *et al.*, 1995). Thus the presence within the U2 repeat unit of both an LTR element and a CT microsatellite may work synergistically to render the U2 tandem array particularly competent for DNA recombination, such as sister chromatid exchange. Another intriguing possibility is that the high concentration of powerful U2 transcription units within the *RNU2* locus interferes with proper chromatin condensation, partially exposing the underlying DNA and causing the locus to be recombinogenic. This could explain why fragile sites are hotspots for sister chromatid exchange (Glover and Stein, 1987) and why the human *RNU2* locus is the major adenovirus 12-induced fragile site (Bailey *et al.*, 1995; Gargano *et al.*, 1995).

Haplotype diversity at the RNU2 locus

We found only two types of U2 tandem arrays in diverse non-African populations, but at least five different types of U2 tandem arrays in African populations based on *Sac*I and CT microsatellite polymorphisms (see Table I and Figure 4). These patterns of haplotype diversity are consistent with genetic evidence suggesting a recent African origin for modern non-African humans (Armour *et al.*, 1996; Tishkoff *et al.*, 1996). Linkage disequilibrium of the two types of U2 tandem arrays among non-African populations (the *Sac*I$^+$ and CT$^+$ polymorphisms are associated, as are the *Sac*I$^-$ and CT$^-$ polymorphisms; Table I) suggests that a limited number of people migrated out of Africa and their descendants populated the rest of the world. Greater haplotype diversity and lack of linkage disequilibrium in the U2 tandem arrays of African populations (independent assortment of the *Sac*I and CT polymorphisms) likewise suggests that the origin of modern humans in Africa substantially predates the emigration out of Africa. The data also underscore the low frequency of interchromosomal recombination among *RNU2* loci; complete linkage disequilibrium was observed in >20 U2 tandem arrays examined from non-African populations, even though length variation within each type of U2 array provides *prima facie* evidence for ongoing genetic activity (Table I).

A model for the mechanism of concerted evolution

To account for our data, we propose a model for concerted evolution of tandemly repeated multigene families (Figure 5). The homogeneity of the polymorphic *Sac*I and CT microsatellite markers within individual U2 tandem arrays suggests that mutations arising within an individual U2 tandem array are eliminated rapidly or spread throughout the array by intrachromosomal recombination processes such as USCE and/or intrachromatid gene conversion. The absence of reciprocal recombination between the dimorphic, tightly associated, flanking haplotypes suggests that slower interallelic genetic exchange between homologous (non-sister) chromosomes occurs by gene conversion. These gene conversion-like events need not be simple; tandem gene organization may allow single repeats or blocks of repeats to be swapped at the same time, as observed for certain human minisatellite loci (Wolff *et al.*, 1989; Buard and Vergnaud, 1994; Jeffreys *et al.*, 1994). Gene conversion may be initiated by double strand breaks (DSBs), as suggested for transposon-mediated conversion (Thompson-Stewart *et al.*, 1994) and minisatellite evolution (Jeffreys *et al.*, 1994) or staggered single-stranded nicks (SSSN), as proposed for complex recombination events at minisatellite loci (Buard and Vergnaud, 1994). Since interchromosomal recombination is thought to be much less frequent than intrachromosomal recombination (Shulman *et al.*, 1995), linkage disequilibrium between markers flanking the U2 tandem array may persist for long periods of time. Following such interchromosomal 'cross-talk' events, additional rounds of rapid intrachromosomal exchange would then homogenize and ultimately fix the mutation in the recipient array. We agree with Schlötterer and Tautz (1994) who concluded, based on studies of *Drosophila* rDNA, that the homogeneity of tandemly repeated genes in metazoans must be maintained by intrachromosomal events; however, our data documenting the absence of reciprocal recombination between flanking markers enable us to conclude, in addition, that new alleles are introduced into the tandem array by interchromosomal gene conversion.

Materials and methods

Preparation of DNA samples

Genomic DNAs were generally isolated as agarose plugs and digested with restriction enzymes within the plugs. Genomic DNAs were prepared from EBV-transformed lymphocyte lines unless otherwise specified. When preparative FIGE was used to recover individual U2 tandem arrays for restriction digestion, the gel was fractionated into slices, and each slice was then treated with β agarase, phenol extracted and the DNA precipitated with ethanol in the presence of carrier nucleic acid (DNA or RNA). Genomic 'unblotting' was carried out as described (Liao and Weiner, 1995). Individual U2 arrays were also isolated from dried agarose gels after unblotting, and the gel slices containing the U2 arrays of interest were melted in TE and a portion used as template for allele-specific PCR amplification essentially as described (Liao and Weiner, 1995).

Array-specific PCR

Array-specific PCR and PCR primers for amplification of the CT microsatellite were as described (Liao and Weiner, 1995). PCR primers for amplification of the junction regions of human U2 tandem arrays were U2JR1 (5′-ACCACTGAAGCACAGCATCA-3′, corresponding to positions –581 to –562 of JR), U2JR2 (5′-TAACAGCGTAGCTA-GCCTTC-3′, complementary to the sequence between +158 and +177 of JR), U2JL1 (5′-AGACTGAGGCATGAGAATCA-3′, corresponding to positions –353 to –334 of JL) and U2JL2 (5′-ACACAGAGTTAGG-AGCTGAA-3′, complementary to nucleotides +241 to +223 of JL). PCR primers used for amplifying the *Sac*I$^{+/-}$ region of a U2 repeat were U2Sac4 (5′-TACTGAGCGCCTTCCACACG-3′, corresponding to nucleotides 3822–3841 of the 6.1 kb U2 repeat) and U2Sac5 (5′-AGACAGAACCGGAAGAGACC-3′, complementary to nucleotides 4543–4524 of the U2 repeat). Coordinates for nucleotide positions were arbitrary and begin at the *Hin*dIII site according to Pavelitz *et al.* (1995); the reported sequence of the U2 repeat (GenBank accession No. L37793) subsequently has been revised (see accession No. U57614).

DNA cloning and sequencing

Gel-purified PCR products were either sequenced directly or cloned in the pGEM-T® vector (Promega) and sequenced. For sequencing PCR fragments directly, a DNA fragment was mixed with a sequencing primer and Sequenase® buffer. The mixture was then boiled in a water bath for 5–10 min, and quickly quenched on ice. Cold labeling mix was added, and the labeling reaction was allowed to continue for 1–5 min on ice before termination. Otherwise, the standard Sequenase® protocol was followed.

Acknowledgements

We thank Cathy Barr for her tireless efforts to find *RNU2* RFLPs, and Russell Bell of Myriad Genetics for the kind gift of P1 genomic clones spanning the human *RNU2* locus. This work was supported by NIH grants GM41624 and GM31073 to A.M.W., NIH grant MH39239 to K.K.K. and a Medical Research Council of Canada Postdoctoral Fellowship awarded to D.L.

References

Armour,J.A.L. *et al*. (1996) Minisatellite diversity supports a recent African origin for modern humans. *Nature Genet*., **13**, 154–160.

Arnheim,N., Krystal,M., Schmickel,R., Wilson,G., Ryder,O. and Zimmer,E. (1980) Molecular evidence for genetic exchanges among ribosomal genes on non-homologous chromosomes in man and apes. *Proc. Natl Acad. Sci. USA*, **77**, 7323–7327.

Bailey,A.D., Li,Z., Pavelitz,T. and Weiner,A.M. (1995) Adenovirus type 12-induced fragility of the human *RNU2* locus requires U2 small nuclear RNA transcriptional regulatory elements. *Mol. Cell. Biol*., **15**, 6246–6255.

Buard,J. and Vergnaud,G. (1994) Complex recombination events at the hypervariable minisatellite CEB1 (D2S90). *EMBO J*., **13**, 3203–3210.

Charlesworth,B., Sniegowski,P. and Stephan,W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, **371**, 215–220.

Coen,E.S. and Dover,G.A. (1983) Unequal exchanges and the coevolution of X and Y rDNA arrays in *D.melanogaster. Cell*, **33**, 849–855.

Dahlberg,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.

Dib,C. *et al*. (1996) A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380**, 152–154.

Dover,G.A. (1982) Molecular drive, a cohesive mode of species evolution. *Nature*, **299**, 111–117.

Dover,G.A., Linares,A.R., Bowen,T. and Hancock,J.M. (1993) Detection and quantification of concerted evolution and molecular drive. *Methods Enzymol*., **224**, 525–541.

Elder,J.F.,Jr and Turner,B.J. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *Q. Rev. Biol*., **70**, 297–320.

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.

Gargano,S., Wang,P., Rusanganwa,E. and Bacchetti,S. (1995) The transcriptionally competent U2 gene is necessary and sufficient for adenovirus type 12 induction of the fragile site at 17q21–22. *Mol. Cell. Biol*., **15**, 6256–6261.

Glover,T.W. and Stein,C.K. (1987) Induction of sister chromatid exchanges at common fragile sites. *Am. J. Hum. Genet*., **41**, 882–890.

Gonzalez,I.L., Sylvester,J.E. and Schmickel,R.D. (1988) Human 28S ribosomal RNA sequence heterogeneity. *Nucleic Acids Res*., **16**, 10213–10224.

Gonzalez,I.L., Wu,S., Li,W.M., Kou,B.A. and Sylvester,J.E. (1992) Human ribosomal RNA intergenic spacer sequence. *Nucleic Acids Res*., **20**, 5846.

Hammarstrøm,K., Westin,G., Bark,C., Zabielski,J. and Pettersson,U. (1984) Genes and pseudogenes for human U2 RNA. Implications for the mechanism of pseudogene formation. *J. Mol. Biol*., **179**, 157–169.

Hibner,B.L., Burke,W.D. and Eickbush,T.H. (1991) Sequence identity in an early chorion multigene family is the result of localized gene conversion. *Genetics*, **128**, 595–606.

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.

Htun,H., Lund,E., Westin,G., Pettersson,U. and Dahlberg,J.E. (1985) Nuclease S1-sensitive sites in multigene families, human U2 small nuclear RNA genes. *EMBO J*., **4**, 1839–1845.

Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature*, **314**, 67–73.

Jeffreys,A.J., Tamaki,K., MacLeod,A., Monckton,D.G., Neil,D.L. and Armour,J.A.L. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet*., **6**, 136–145.

Jinks-Robertson,S. and Petes,T.D. (1993) Experimental determination of rates of concerted evolution. *Methods Enzymol*., **224**, 631–646.

Kadyk,L.C. and Hartwell,L.H. (1992) Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics*, **132**, 387–402.

Kadyk,L.C. and Hartwell,L.H. (1993) Replication-dependent sister chromatid recombination in *rad1* mutants of *Saccharomyces cerevisiae*. *Genetics*, **133**, 469–487.

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.

Lindgren,V., Ares,M., Weiner,A.M. and Francke,U. (1985) Human genes for U2 small nuclear RNA map to a major adenovirus 12 modification site on chromosome 17. *Nature*, **314**, 115–116.

Lovett,S.T., Drapkin,P.T., Sutera,V.A.,Jr and Gluckman-Peskind,T.J. (1993) A sister-strand exchange mechanism for recA-independent deletion of repeated DNA sequences in *Escherichia coli*. *Genetics*, **135**, 431–642.

Mangin,M., Ares,M.,Jr and Weiner,A.M. (1985) U1 small nuclear RNA genes are subject to dosage compensation in mouse cells. *Science*, **229**, 272–275.

Matera,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.

Mellon,I., Rajpal,D.K., Koi,M., Boland,C.R. and Champe,G.N. (1996) Transcription-coupled repair deficiency and mutations in human mismatch repair genes. *Science*, **272**, 557–560.

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.

Murti,J.A., Bumbulis,M. and Schimenti,J.C. (1994) Gene conversion between unlinked sequences in the germline of mice. *Genetics*, **137**, 837–843.

Nagele,R., Freeman,T., McMorrow,L. and Lee,H.-y. (1995) Precise spatial positioning of chromosomes during prometaphase, evidence for chromosome order. *Science*, **270**, 1831–1835.
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.
Petes,T.D. (1980) Molecular genetics of yeast. *Annu. Rev. Biochem*., **49**, 845–876.
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. (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res*., **20**, 211–215.
Schlötterer,C. and Tautz,D. (1994) Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr. Biol*., **4**, 777–783.
Seperack,P., Slatkin,M. and Arnheim,N. (1988) Linkage disequilibrium in human ribosomal genes, implications for multigene family evolution. *Genetics*, **119**, 943–949.
Shiroishi,T., Koide,T., Yoshini,M., Sagai,T. and Moriwaki,K. (1995) Hotspots of homologous recombination in mouse meiosis. *Adv. Biophys*., **31**, 119–132.
Shulman,M.J., Collins,C., Connor,A., Read,L.R. and Baker,M.D. (1995) Interchromosomal recombination is suppressed in mammalian somatic cells. *EMBO J*., **14**, 4102–4107.
Smith,G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science*, **191**, 528–535.
Sorensen,P.D. and Frederiksen,S. (1991) Characterization of human 5S rRNA genes. *Nucleic Acids Res*., **19**, 4147–4151.
Szostak,J.W. and Wu,R. (1980) Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature*, **284**, 426–430.
Thompson-Stewart,D., Karpen,G.H. and Spradling,A.C. (1994) A transposable element can drive the concerted evolution of tandemly repetitious DNA. *Proc. Natl Acad. Sci. USA*, **91**, 9042–9046.
Tishkoff,S.A. *et al*. (1996) Global patterns of linkage disequilibrium at the *CD4* locus and modern human origins. *Science*, **271**, 1380–1387.
Van Arsdell,S.W. and Weiner,A.M. (1984) Human genes for U2 small nuclear RNA are tandemly repeated. *Mol. Cell. Biol*., **4**, 492–499.
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.
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.
Willard,H.F. (1990) Centromeres of mammalian chromosomes. *Trends Genet*., **6**, 410–416.
Wolff,R.K., Plaetke,R., Jeffreys,A.J. and White,R. (1989) Unequal crossingover between homologous chromosomes is not the major mechanism involved in the generation of new alleles at *VNTR* loci. *Genomics*, **5**, 382–384.

Received on August 13, 1996; revised on September 24, 1996