Evolution of Tandemly Repeated Sequences: What Happens at the End of an Array?

Bryant F. McAllister, John H. Werren

Department of Biology, University of Rochester, Rochester, New York, USA

Abstract. Tandemly repeated sequences are a major component of the eukaryotic genome. Although the general characteristics of tandem repeats have been well documented, the processes involved in their origin and maintenance remain unknown. In this study, a region on the paternal sex ratio (PSR) chromosome was analyzed to investigate the mechanisms of tandem repeat evolution. The region contains a junction between a tandem array of PSR2 repeats and a copy of the retrotransposon *NATE,* with other dispersed repeats (putative mobile elements) on the other side of the element. Little similarity was detected between the sequence of PSR2 and the region of *NATE* flanking the array, indicating that the PSR2 repeat did not originate from the underlying *NATE* sequence. However, a short region of sequence similarity (11/15 bp) and an inverted region of sequence identity (8 bp) are present on either side of the junction. These short sequences may have facilitated nonhomologous recombination between *NATE* and PSR2, resulting in the formation of the junction. Adjacent to the junction, the three most terminal repeats in the PSR2 array exhibited a higher sequence divergence relative to internal repeats, which is consistent with a theoretical prediction of the unequal exchange model for tandem repeat evolution. Other *NATE* insertion sites were characterized which show proximity to both tandem repeats and complex DNAs containing additional dispersed repeats. An "accretion model" is proposed to account for this association by the accumulation of mobile elements at the ends of tandem arrays and into "islands" within arrays. Mobile elements inserting into arrays will tend to migrate into islands and to array ends, due to the turnover in the number of intervening repeats.

Key words: Tandemly repetitive DNA — Satellite DNA — Retrotransposons — Junction site — Unequal sister chromatid exchange — Recombination

Introduction

Tandemly repeated sequences, or satellite DNAs, are present in a wide range of organisms, and in certain cases these sequences may constitute a large percentage of the total genome. Because tandemly repeated sequences are so abundant and appear to have little or no functional significance, they are commonly regarded as "selfish" or "junk" DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). Studies of tandem repeats have generally focused on determining the nucleotide sequence, abundance, genomic organization, and/or species distribution of the repeats (Willard 1989). Tandem repeats are routinely divided into distinct classes based on the size of the repeating sequence; such as microsatellite, minisatellite, and satellite DNA (Charlesworth et al. 1994). It is the "classic" satellite sequences (approximately 200 bases in length) which are the focus of this investigation. Two important questions regarding the evolution of satellite sequences remain unresolved: What leads to the origins of these repeats? and What are the dynamics of

Correspondence to: Bryant McAllister at Department of Biology, University of Texas at Arlington, Box 19498, Arlington, TX 76019-0498, USA

repetitive arrays? Many theoretical models have provided a framework in which to address these issues, although the empirical data have been inadequate for distinguishing among the proposed mechanisms underlying repeat evolution.

The first issue is the origins of repetitive arrays. Sequence duplication by an aberrant recombination (Krüger and Vogel 1975; Ohta 1980) or replication slippage (Levinson and Gutman 1987; Walsh 1987; Stephan 1989) may represent the primary event which creates a repeat. Both of these processes involve duplicating a previously single-copy sequence, thus initializing a tandem repeat. After amplification of the sequence into a large array, the final sequence of the repeating unit should resemble the original nonrepetitive sequence. Smith (1976) envisioned a different process of repeat origin. He postulated that a nonfunctional sequence will experience unequal exchange events (mismatches out of register from complete identity), and the recombinant products would not be eliminated by purifying selection. Following a large number of these unequal exchanges, a periodicity evolves in the sequence (Smith 1976; Stephan 1989). This model predicts that repeats originate from an existing sequence, although the relationship between the original sequence and the final repeated unit is indeterminate. The duplication, replication slippage, and unequal exchange models all propose a de novo origin of each repetitive array, but an alternative is that new repetitive arrays evolve from existing arrays at other locations in the genome. Excision and circularization of a few repeats from a tandem array may provide a template for rolling circle replication (Walsh 1987), and following amplification into a linear array, the repeats may insert at a new location in the genome. By examining the sequences immediately flanking repetitive arrays, or by inferring the ancestral state of the underlying single-copy sequence prior to repeat origin, it may be possible to identify patterns that are consistent with one of these mechanisms.

A second issue concerns the dynamics and maintenance of existing arrays. The most widely accepted mechanisms underlying maintenance of arrays of satellite DNA are unequal exchange and intrastrand exchange (Charlesworth et al. 1994). Successive unequal exchanges between paralogous units of repetitive arrays on separate DNA strands prevents individual repeats from evolving independently, instead the arrays evolve in concert (Smith 1976; Ohta 1980; Stephan 1989). Unequal exchanges increase and decrease the number of repeats in an array, which over successive generations reduces sequence variation and causes turnover among repeats by stochastic fixation of a single repeat type. According to this model, homogenization by unequal exchange occurs in regions of the array that are free to recombine; however, unique sequences flanking a repetitive array inhibit

exchanges involving repeats located near the termini (Stephan 1989). These terminal repeats are also influenced by overall rates of unequal exchange, which causes identity among repeats to decrease as spatial distance increases (Ohta 1980). Therefore, evolution by unequal exchange should lead to terminal repeats containing the highest sequence divergence relative to the common repeated sequence. Models of unequal exchange also indicate that higher order organization of repeats can occur (Stephan 1989). This process may be facilitated by the insertion of mobile elements into tandem arrays, although there have been few studies of the pattern and distribution of elements within tandem arrays and little is known about the structure of junctions between tandem arrays and other DNAs.

The Paternal Sex Ratio (PSR) chromosome in the parasitic wasp *Nasonia vitripennis* provides a unique system for studying the origin, maintenance, and structure of tandemly repeated sequences. A key feature of PSR is that it never experiences meiosis and is maintained strictly as a haploid through mitotic cell cycles. PSR is continually maintained in a haploid cell lineage because it is always carried by (haploid) male wasps. The chromosome is transmitted in sperm, and after fertilization PSR causes loss of the paternal autosomes in the first zygotic division (Werren et al. 1987; Nur et al. 1988). The result is development of a male wasp containing the haploid maternal autosomes and PSR. Since spermatogenesis is entirely mitotic in male wasps, PSR never undergoes meiotic recombination. Therefore, all processes involved in the maintenance of tandem repeats on PSR act exclusively during mitotic cell divisions.

The PSR chromosome contains four major families of tandemly repeated sequences (Eickbush et al. 1992). Three of these tandem repeats (PSR2, PSR18, and PSR22) are classic ''satellite'' DNAs and they apparently evolved specifically on PSR, because they are not present on the autosomes of *N. vitripennis* or related species (Nur et al. 1988; Eickbush et al. 1992). The three repeated sequences are approximately 171–214 bp in length (Eickbush et al. 1992) and are present in large localized arrays on the PSR chromosome (Beukeboom and Werren 1993). Overall sequence similarity is evident and two palindromes are highly conserved among the sequences, thus it appears that these repetitive sequences are evolutionarily related (Eickbush et al. 1992; Reed et al. 1994). In this study, junction regions between tandem arrays and regions of nonarray DNA are analyzed to evaluate mechanisms of tandem array evolution. Prompted by the frequent association between tandem repeats and dispersed repeats on the PSR chromosome, a model is also proposed for the accumulation of mobile DNA in ''islands'' and at the ends of arrays.
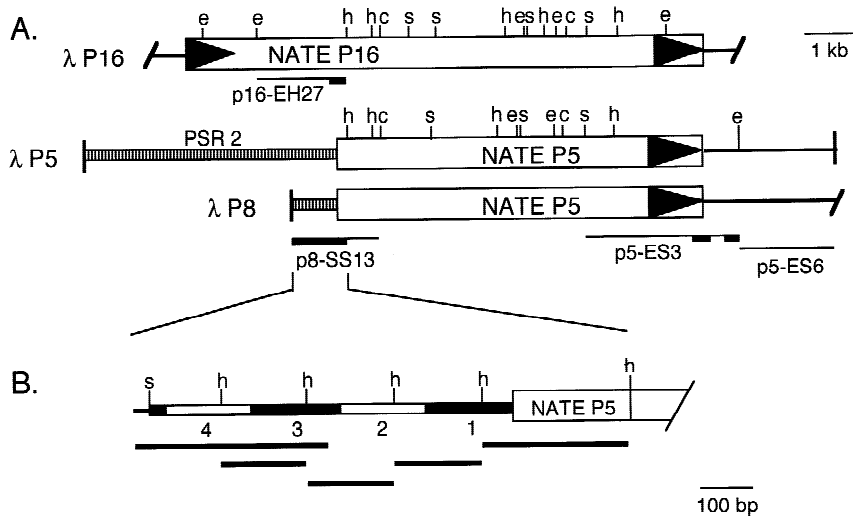
Fig. 1. Structure of clones examined for the junction analysis. **A** The clone λP16 contains a full-length copy of the retroelement *NATE* P16. Both λP5 and λP8 contain the truncated copy of the retroelement *NATE* P5 adjacent to the PSR2 array. *Horizontal lines* below the phage represent subcloned regions, with *heavy lines* indicating sequenced regions. **B** The subcloning and sequencing strategy employed to analyze the PSR2 array. Restriction sites are *Sau*3AI (s) and *Hin*dIII (h) and the *horizontal lines* represent the regions sequenced.

Materials and Methods

The methods for isolating and restriction mapping the λ-clones used in this study have been presented previously (McAllister 1995). Clones were isolated from a genomic library (EMBL 3) constructed from *Sau*3AI partially digested DNA from a standard PSR lab strain (PSR A) crossed to the high fertilizing MI strain of *N. vitripennis* (Reed et al. 1994). Regions of the λ-clones were subcloned into plasmid (Bluescript) vectors for use as probes in Southern blots and as sequencing template. Southern blotting and hybridization followed standard procedures (McAllister 1995; McAllister and Werren 1997).

Characterization of the NATE *P5/PSR2 Junction.* Two λ clones were used for a detailed analysis of a junction region (Fig. 1A); λP8 contains the junction between a truncated copy of the retroelement *NATE* and an array of PSR2 repeats, and λP16 contains a full-length copy of *NATE* and flanking sequences at both ends of the element. Because the PSR2 sequences are organized in tandem, a cloning and sequencing strategy was used whereby the location of each PSR2 repeat in λP8 was identified relative to the junction site (Fig. 1B). The entire array of PSR2 repeats and the adjacent region of *NATE* were subcloned from λP8 by digesting the clone with *Sac*I and *Sal*I and ligating fragments into Bluescript (Stratagene) plasmid, digested with the same enzymes and dephosphorylated with calf intestinal alkaline phosphatase. A clone with an insert of 1.7 kb containing the junction site was obtained and designated p8-SS13 (Fig. 1B). This clone was further subcloned using *Hin*dIII. A 300-bp *Hin*dIII fragment containing the *NATE* P5/PSR2 junction and *Hin*dIII fragments of PSR2 repeat size (171 bp) were subcloned from p8-SS13 into Bluescript (Fig. 1B).

The junction site, including 63 bases of the first PSR2 repeat, was sequenced using the 300-bp *Hin*dIII fragment as template and sequencing both strands. Sequencing of the PSR2 repeats at the other end of the cloned region was initiated by reading one strand of p8-SS13 from the cloning site (of λP8) into the PSR2 array about 370 bases (Fig. 1B). This sequence spanned 135 bases of the PSR2 array before encountering the first *Hin*dIII site, one complete stretch of 171 bases flanked by *Hin*dIII sites, and 65 bases into the adjacent *Hin*dIII fragment. Six clones containing *Hin*dIII fragments of PSR2 repeat size were fully sequenced on both strands. Three clones were identical to the 171-base *Hin*dIII fragment sequenced from p8-SS13, two shared unique nucleotide substitutions with the 65-base incomplete *Hin*dIII fragment, and one contained unique nucleotide substitutions that had not been identified in other *Hin*dIII fragments. Upon organization of these sequences into a linear array, three *Hin*dIII fragments of PSR2 repeat size were

present; this was consistent with mapping estimates of λP8 and p8-SS13 indicating the presence of three *Hin*dIII fragments containing PSR2 repeats.

Sequences from the reverse transcriptase (RT) domain from multiple copies of *NATE* cloned from the PSR chromosome revealed a close relationship among the elements (McAllister 1995). The RT sequence of the element at this junction (*NATE* P5) shares 99.0% (597 bp compared) identity with the same region of the full-length element *NATE* P16. Because these elements appear to have recently replicated from a common ancestor, the region corresponding to the junction in *NATE* P5 was subcloned from *NATE* P16 to obtain the sequence of this region in a ''normal'' copy of *NATE*. The enzymes *Eco*RI and *Hin*dIII were used to subclone λP16 into Bluescript (Fig. 1A). A clone with a 1.7-kb insert homologous to the junction region in *NATE* P5 was obtained and this clone was used for sequencing.

All sequencing reactions were performed on chemically (NaOH) denatured double-stranded plasmid DNA (Ausubel et al. 1992), using the Sequenase (USB) kit and labeling with $[\alpha^{35}S]$ dATP. Reaction products were electrophoresed through buffer gradient gels. Once sequences were obtained, manipulations were performed using ESEE (Cabot and Beckenbach 1989). Dot-plot comparisons were performed using the algorithm of Maizel and Lenk (1981) available in GCG (Genetics Computer Group, Madison, WI, Version 8.0, 1994).

Amplification of the NATE *P5/PSR2 Junction.* To verify the presence of the junction between PSR2 and *NATE* P5 on the PSR chromosome and to determine if the junction site is variable among chromosomes, primers flanking the junction were used to amplify the region from five different isolates of PSR using the polymerase chain reaction (PCR). Template DNA was extracted from males carrying PSR A, the standard laboratory strain that was used to make the λ library, and from four additional isolates of the PSR chromosome. All of these strains were collected from natural populations of *N. vitripennis* in northeastern Utah. The region was amplified using a primer which anneals to the PSR2 repeat (PSR2R, GCT TCT TCA TTT AAG ACG) and another that anneals to *NATE* (N5, CCA TCC TGT GCA GCC TAG). Reactions volumes were 50 μl; containing 1/10 reaction buffer (BRL), 2.5 m*M* MgCl$_2$, a 200 n*M* concentration of each primer, a 100 n*M* concentration of each dNTP, 2 U of *Taq* polymerase (BRL), and approximately 5 ng of template DNA. A Perkin Elmer Cetus DNA thermal cycler was used with the following parameters: 2× (95°C, 2 min; 54°C, 1 min; 72°C, 4 min) linked to 35× (95°C, 0.5 min, 57°C, 1 min, 72°C, 2 min). Reaction products were digested with *Hin*dIII and electrophoresed through a 1.5% agarose gel to determine the size of the *Hin*dIII fragment containing the junction. The products were also cloned into

A.
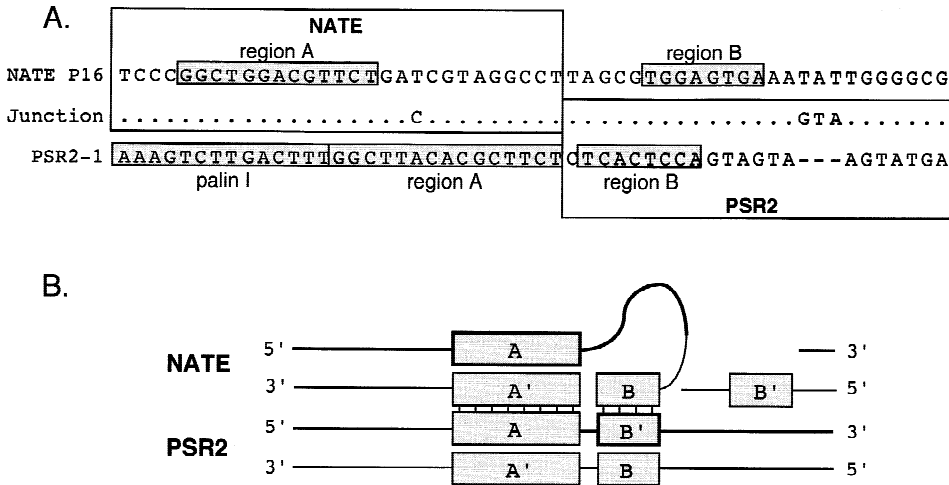
Fig. 2. **A** Sequence of the junction between *NATE* P5 and PSR2. In each *large boxed region,* the junction sequence is compared to either reference sequence of *NATE* P16 or PSR2-1, with *dots* indicating that the same nucleotide is present. Two stretches of sequence similarity, region A and region B, and the palindrome I of PSR2, are indicated by *shaded boxes.* **B** Proposed pairing configuration involving the regions of similarity that could have promoted nonhomologous recombination between *NATE* and PSR2. The *heavier lines* indicate the observed junction resulting from nonhomologous recombination.

Bluescript using *Hin*dIII, and the sequence was obtained from both strands of one clone containing the junction site from each chromosome.

Results

Several regions containing different copies of the retrotransposon *NATE* have been cloned from the PSR chromosome (McAllister 1995). These are described later. First, we describe a detailed analysis of a junction between a copy of *NATE* and an array of PSR2 repeats. The region reveals patterns of repeat evolution and nonhomologous recombination.

Analysis of the NATE/*PSR2 Junction*

General features of the *NATE*/PSR2 junction are shown in Fig. 1, based on two λ clones containing this region. Restriction digestion patterns are identical in the 11-kb region shared by the two clones, thus the inserts apparently represent overlapping frames of the same region. Hybridization to restriction-digested phage DNA revealed that the PSR2 array localizes to one end of *NATE* P5, and upon comparisons to intact copies of *NATE,* truncation of this end (5′) of the element is evident (Fig. 1A). The other end (3′) of *NATE* P5 appears intact (Fig. 1A). To confirm the hybridization data, a 3.0-kb fragment encompassing the LTR, insertion site, and flanking sequence was subcloned from λP5 (p5-ES3 in Fig. 1A). A primer homologous to a site within the LTR was used to read the terminal sequence of the LTR and 100 bases flanking the element. Consistent with the hybridization data, the 3′ end of the element is intact compared to LTR sequences of other elements (data not shown).

Neither the sequence immediately flanking the intact end of *NATE* P5 or a 150-bp sequence located about 1 kb away from the insertion site exhibited any similarity with the standard PSR2-1 sequence upon visual comparison and by dot-plot analysis. Within these flanking sequences, no repetition was identified. The sequence flanking this element was very AT rich. Overall, the sequence was 72% AT, and many polynucleotide stretches of A and T were observed. Southern hybridization revealed that this region contains a middle-repetitive sequence present on both the PSR chromosome and the standard chromosomal complement (data not shown). Thus, the region flanking *NATE* to one side contains PSR2 repeats, whereas the other flanking region does not.

In both clones containing the junction, one cloning site fell within the PSR2 array (Fig. 1A). A restriction map indicated that one clone (λP8) contains approximately four PSR2 repeats. This small number of repeats made λP8 amenable for a detailed study of the junction region, and it was subcloned for further analysis (Fig. 1B).

An approximately 300-bp region of the PSR chromosome encompassing the *NATE* P5/PSR2 junction was subcloned and sequenced. Sequence was obtained for 227 bases of *NATE* P5 leading up to the junction site and for 63 bases of the first PSR2 repeat. The homologous region, and an additional 100 bases representing what would be on the other side of the junction site (if *NATE* P5 was not truncated at the junction), was sequenced from *NATE* P16. Figure 2A shows the junction sequence compared to sequences from *NATE* P16 and a reference PSR2 repeat [PSR2-1 (Eickbush et al. 1992)]. This comparison reveals an abrupt transition between the *NATE* and the PSR2 sequences, with only a ''CT'' being shared

at the junction site. The copy of *NATE* at the junction exhibits 96.9% sequence identity with the reference element throughout 227 bp leading up to the junction site, where there is an immediate transition to a PSR2 sequence with 92.1% (over 63 bp excluding a 3-bp indel) identity with the reference repeat.

Origin of the NATE/*PSR2 Junction*

As shown in Fig. 2A, two short stretches of sequence similarity between *NATE* and PSR2 are present on either side of the junction site (designated A and B). Region A is a stretch of 11 bases of identity shared between a 13-base sequence in *NATE* and a 15-base sequence in PSR2. Region A is 12 bp upstream of the junction in *NATE* and immediately upstream of the junction in PSR2. Region B is 1 base downstream of the junction in PSR2 and 5 bases downstream in *NATE.* It consists of eight bases in *NATE,* with the exact complement in reverse orientation present in PSR2. Short regions of sequence similarity are known to be associated with recombination events (Metzenberg et al. 1991; Warburton et al. 1993; Sakagami et al. 1994). This region is unusual in having two short stretches with one a compliment in reverse orientation. It seems unlikely that these two regions of notable identity occur in such close proximity of the junction by chance.

A possible configuration leading to recombination between *NATE* and PSR2 is presented in Fig. 2B. This configuration involves illegitimate homologous pairing between *NATE* and PSR2, either within the same chromosome or between sister chromatids, and would have occurred mitotically (due to the paternal inheritance pattern of PSR). Assuming some sort of DNA damaging event, such as a single-strand or double-strand break, illegitimate homologous pairing between *NATE* and PSR2 could have occurred in region A, with this short region of homology being further stabilized by *NATE* region B pairing with the complementary strand of PSR2 region B (Fig. 2B). The formation of the hybrid structure with region B is made possible by the looping of the 17-base stretch separating A and B within *NATE.* In addition, a short palindromic region (AGGCCT) may have further facilitated the DNA bending that was required. Resolution of the configuration presumably led to a break between A and B, with recombination between *NATE* and PSR2 at this site causing truncation of the *NATE* element. It is also interesting to note that this region of the PSR2 sequence has been involved in recombination with another tandemly repetitive sequence (Reed et al. 1994) and is adjacent to a 14-bp palindromic region (palin I) present in the PSR2 repeat (Fig. 2A). Palindromes are known to facilitate DNA melting (Chalker et al. 1993) and to be associated with recombination events (Bigot et al. 1990; Reed et al. 1994).

We investigated several alternatives to nonhomolo-gous recombination that may have led to the formation of this junction. To determine if there is evidence that the PSR2 array is encroaching on *NATE,* the junction site was determined on different chromosomes. The junction site was amplified by PCR, cloned, and sequenced from PSR A (lab standard used to construct λ library) and four additional PSR chromosomes that were collected independently from natural populations. Identical junction sites between *NATE* P5 and PSR2 were found on all five PSR chromosomes, confirming that the junction site is not a cloning artifact and indicating its stability during the time interval since the divergence of these chromosomes.

Another alternative to nonhomologous recombination is that the PSR2 repeat family evolved from *NATE.* To determine if there were large-scale similarities between the PSR2 sequence and the region of *NATE* underlying the junction, dot-plot comparisons (Maizel and Lenk 1981) were performed. A 229-bp region of *NATE* P5 leading up to the junction site, the four PSR2 repeats, and 100 bp of *NATE* P16 on the other side of the junction site were joined into a composite sequence and this was compared to the PSR2-1 sequence. Window sizes of 8 and 21 bases were used, with stringencies of 50, 75, and 100%. Similarity among the four PSR2 repeats was evident in all of the dot-plot analyses, although no remarkable similarity was detected between the PSR2 repeat and *NATE.* Therefore, except for the two short regions of similarity directly adjacent to the junction, there is no significant similarity between *NATE* and PSR2. The observed pattern is more consistent with a recombination event between *NATE* and PSR2 than with PSR2 arising from *NATE.*

Another possible explanation for the formation of this junction is an insertion of PSR2 repeats into *NATE.* If the PSR2 array inserted into *NATE,* it is expected that the remaining piece of *NATE* P5 should be flanking the other end of this PSR2 array. Both PCR assays and a PSR library screen were used in attempts at isolating the expected junction at the other end of the array. A combination of primers were used in an attempt to amplify the expected junction. Two primers were used that were located within the LTR of *NATE* and oriented in the same direction (N8, 5′ ACG ATT ACA TAA AGC ATA G; N7, 5′ CAC GAG GCA TCA TCT GCG), in combination with primers homologous to the PSR2 and PSR18 sequences (2R, 5′ GGC CTC AGG ACG GCG TTC; 2F, 5′ GAA CGC CGT CCT GAG GCC; 18R, CGA CAT CTT TTC CTC AGC G; 18F, 5′ GAG GAA AAG ATG TCG CTC C). Based on the organization of the PSR2 sequence at the *NATE* P5/PSR2 junction, either *NATE* primer in combination with the 2R primer should have produced the expected product. Attempts at amplifying the junction with all of these primer combinations were unsuccessful (data not shown). In addition, the PSR genomic library was screened to determine if clones rep-
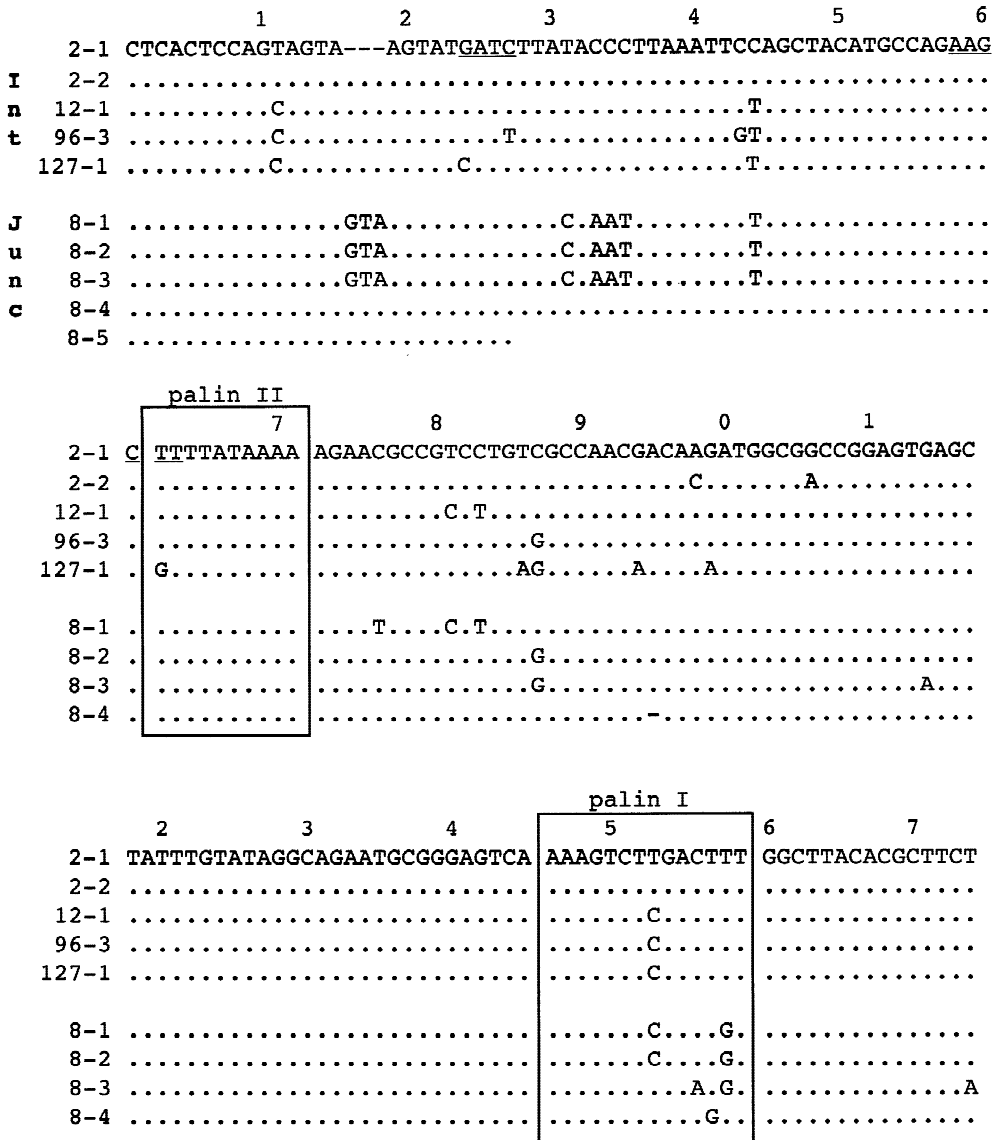
```
                1            2            3            4            5            6
     2-1  CTCACTCCAGTAGTA---AGTATGATCTTATACCCTTAAATTCCAGCTACATGCCAGAAG
I    2-2  ............................................................
n   12-1  .........C....................................T.............
t   96-3  .........C..................T.................GT.............
   127-1  .........C...................C...................T..........

J    8-1  ..............GTA............C.AAT........T.................
u    8-2  ..............GTA............C.AAT........T.................
n    8-3  ..............GTA............C.AAT........T.................
c    8-4  ............................................................
     8-5  ..........................

            palin II
                          7           8           9           0           1
     2-1  C TTTTATAAAA AGAACGCCGTCCTGTCGCCAACGACAAGATGGCGGCCGGAGTGAGC
     2-2  . .......... ..............................C.......A..........
    12-1  . .......... ........C.T....................................
    96-3  . .......... ...............G...............................
   127-1  . G......... ............AG......A....A......................

     8-1  . .......... ....T....C.T.................................
     8-2  . .......... ...............G...............................
     8-3  . .......... ...............G...........................A...
     8-4  . .......... ................-..............................

                                        palin I
            2            3            4           5            6            7
     2-1  TATTTGTATAGGCAGAATGCGGGAGTCA AAAGTCTTGACTTT GGCTTACACGCTTCT
     2-2  ..........................  ..............  ..............
    12-1  ..........................  .......C......  ..............
    96-3  ..........................  .......C......  ..............
   127-1  ..........................  .......C......  ..............

     8-1  ..........................  .......C....G.  ..............
     8-2  ..........................  .......C....G.  ..............
     8-3  ..........................  ..........A.G.  ..............A
     8-4  ..........................  ..........G..   ..............
```

Fig. 3. Sequences of five previously published PSR2 sequences that represent internal repeats and four complete and one partial repeats obtained from the junction region. Recognition sites for *Sau*3AI and *Hin*dIII are *underlined,* and the palindromic regions are indicated.

resenting the expected structure were present. The region of *NATE* P16 (p16-EH27) that encompasses the junction in *NATE* P5 was used as a probe to screen about five genome equivalents of the library. Approximately 150 clones hybridized strongly to the probe. Thirty-two clones were isolated, regrown, and screened for hybridization to the PSR2 repeat. No clones were identified which hybridized to both probes. Therefore, these experiments failed to identify the expected organization at the other end of the array.

Sequences of PSR2 Repeats at an Array Terminus

To investigate the evolution of repeats at the junction region, the number and linear structure of PSR2 repeats in this region was determined. The observed pattern is consistent with unequal exchange being inhibited at the

junction with *NATE*. After the sequence of the PSR2 array present in λP8 was obtained, the presence of the three *Hin*dIII restriction fragments of PSR2 was verified. Using the initiation of the first PSR2 repeat to define the periodicity in the array, four complete PSR2 repeats (approx 171 bp each) and one partial PSR2 repeat (27 bp) were identified in this sequence (Fig. 3). Even though the entire sequence of the array was not obtained in a linear fashion, by sequencing overlapping clones the relative positioning of the PSR-2 repeats is this region was determined (see Materials and Methods). The repeat which forms the junction with *NATE* P5 is referred to as 8-1, and the others are numbered consecutively (8-2, etc.) into the array (Fig. 3).

For comparative purposes, five ''random'' PSR2 repeat sequences were available (Eickbush et al. 1992). These published PSR2 repeats presumably represent a
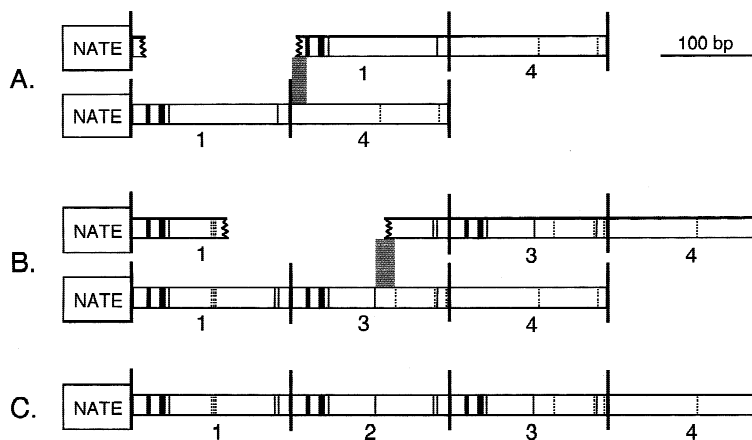
Fig. 4. Inferred history of unequal exchange between repeats adjacent to the junction with *NATE*. Nucleotide sites that are present in at least two PSR2 repeats (but not all) are represented as *solid lines,* whereas the sites only present in a single repeat are represented by *dotted lines.* Recombination events are depicted as one chromatid with a double-strand break using the other chromatid as template for repair. The *shaded boxes* indicate the window where the exchange event is localized for the strand containing the PSR 2 repeats. **A** Oldest inferred unequal exchange shifting the block of unique nucleotides into the array. **B** Unequal exchange creating the mosaic second repeat. **C** Structure of the array in its present form.

random sample of repeats located internally within the array, because they were subcloned and sequenced from four λ clones, each with 15- to 20-kb inserts consisting entirely of PSR2 repeats. PSR2 sequences on the PSR chromosome are remarkably homogeneous in restriction enzyme digestion patterns, there is no evidence of a higher-order structure in the array, and they are localized in one or a few arrays (Eickbush et al. 1992; Beukeboom and Werren 1993; J.H.W., unpublished data). As is evident in the visual comparison of repeats in Fig. 3, the PSR2 repeats adjacent to the junction site are similar to the internal PSR2 repeats. Nevertheless, there are a number of unique nucleotides in the PSR2 repeats at the end of this array which have not been identified in the other PSR2 repeats (Fig. 3).

Sequence comparisons were performed to quantify the differences between the PSR2 repeats at the junction relative to the internal repeats. The mean pairwise sequence difference among the five internal repeats was 0.044 ± 0.019 (SD). Each of the four repeats at the junction was compared to the five internal repeats, and means and standard deviations of these pairwise comparisons were calculated. The sequence difference between the first, second, and third repeats was 0.068 ± 0.021, 0.056 ± 0.012, and 0.075 ± 0.015, respectively, which was higher than the calculated difference among the internal repeats. The fourth repeat in the array 8-4 had a mean difference of 0.034 ± 0.019, lower than the value observed among the internal repeats. These sequence comparisons reveal a trend where the first three repeats in the array are more divergent from the internal repeats than the internal repeats are to each other, and the fourth repeat is very similar to the internal repeats.

The distance estimates indicate that the three most terminal repeats are distinct from other PSR2 repeats, although visual comparisons reveal many of the nucleotide positions which distinguish these first three repeats are clustered and shared among them (Fig. 3). In contrast to the first three repeats, the fourth repeat does not contain these distinct features and it is almost identical to the internal PSR2 repeats. Although the first three repeats

are similar, different regions of each repeat may represent units with their own history, and unequal exchange events have brought these units together into the extant repeats which were observed. One apparent example of this pattern is in repeat 8-2 from the PSR2 array terminus. This repeat is a perfect mosaic containing partial regions of the first and third repeats. By postulating an array containing the first and third repeats, a single unequal exchange between these two repeats would generate the second repeat (Fig. 4). This unequal exchange event appears to be recent, because no nucleotide substitutions are apparent in the 348 bases representing the mosaic second repeat compared to the progenitor segments. In addition to the apparent mosaic nature of the second repeat, a block of distinctive nucleotides that have not been identified in other PSR2 repeats are shared by the first, second, and third repeats (Fig. 3 and 4). The presence of these nucleotides in the second repeat is consistent with the previously mentioned unequal exchange, but their presence in the third repeat indicates another unequal exchange. These distinctive nucleotides may have been present at the formation of the junction, with subsequent divergence of the more interior array, or the terminal repeats may have diverged from the internal array by mutations following formation of the junction. In either case, it appears that unequal exchanges have moved this segment into the array. Furthermore, at least three nucleotide substitutions have occurred in both the first and the third repeats during the time separating these two unequal exchanges. The pattern observed indicates that the repeats adjacent to *NATE* are divergent from consensus PSR2 repeats but converge on the consensus sequence by the fourth repeat.

General Structure of *NATE-Repeat Junctions*

The structure of five additional regions containing *NATE* inserts are shown in Fig. 5. The results are based on cloning, restriction digestion analyses, and Southern hybridizations using subclones from these regions as
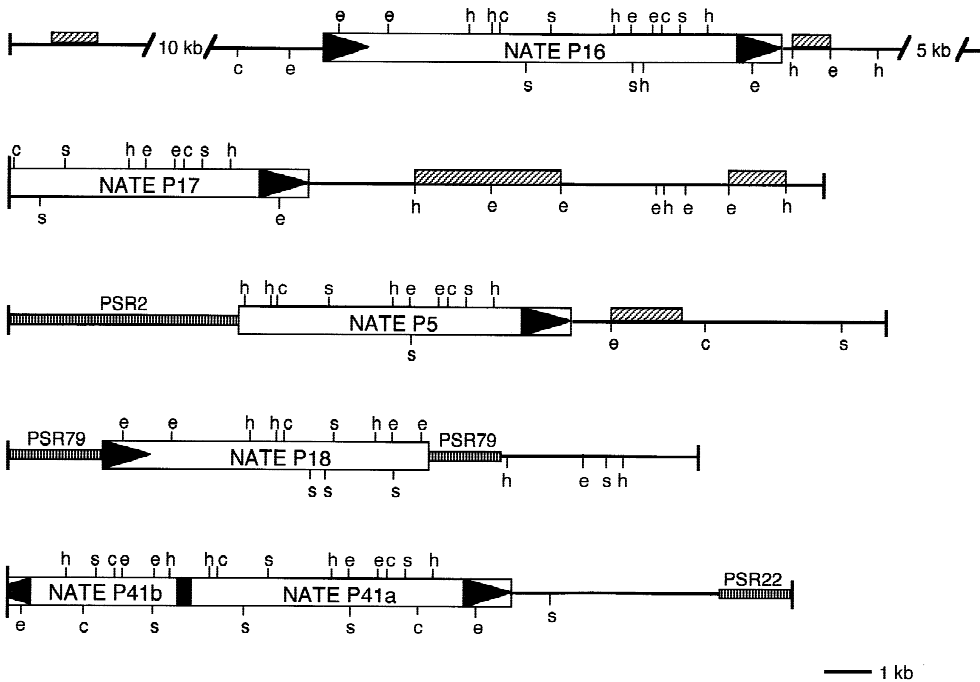
Fig. 5. Cloned regions from the PSR chromosome. Positions of the retroelement *NATE* are indicated by the *large rectangles,* with the *black arrowheads* representing the LTRs. The *blocks of vertical lines* represent arrays of the tandemly repetitive sequences PSR2, PSR22, and PSR79. Regions flanking the copies of *NATE* which have middle-repetitive hybridization patterns are represented by *cross-hatched boxes.*

probes to restriction-digested genomic DNA. As a general pattern, *NATE* inserts are found near tandem arrays and near other dispersed repetitive sequences. Three of the five junctions contain nearby (or immediately flanking) regions of known tandemly repetitive DNA. As described previously, *NATE* P5 has a junction with an array of PSR2, with apparently dispersed repetitive DNA flanking the other end of the element. *NATE* P18 is a truncated retroelement with PSR79 on one side and a short (about 1.5-kb) stretch of PSR79 on the other, followed by a complex nonarray region. *NATE* P41b/41a are adjacent truncated elements, with approximately 5 kb of complex DNA followed by a region of PSR22 repeats. The other two clones contain *NATE* elements adjacent to complex non-array DNA. The regions of complex DNA flanking these elements are primarily composed of a diverse set of dispersed repetitive DNA (for an example, see McAllister and Werren 1997). A high frequency of truncation is also observed among these copies of *NATE;* of six elements, at least four are truncated.

Several specific features are noteworthy (Fig. 5). *NATE* P41a and b are two back-to-back truncated element in the vicinity of a PSR22 array. It is estimated that there are nine copies of *NATE* on the 40-Mb PSR chromosome. The probability of these two elements being associated by chance is very low. They must have either (a) preferentially inserted next to each other or (b) inserted at more distant location and subsequently "moved" into close proximity. This may have been facilitated by nonhomologous recombination. *NATE* P18 also shows interesting features. It is very close to the end

of a PSR79 array. PSR79 is a complex tandemly repeated sequence (Eickbush et al. 1992), and current deletion studies indicate that it occurs in a localized block on the PSR chromosome (Beukeboom and Werren 1993; McAllister, Beukeboom, and Werren, in preparation). The probability that *NATE* P18 inserted at the end of a PSR79 array is thus low, suggesting migration of this element toward the end of an array after insertion. *NATE* P5 is at the end of a PSR2 array and near other mobile elements. Both *NATE* P17 and *NATE* P16 occur in complex genetic regions containing other dispersed repetitive DNA (presumed mobile elements). The general pattern suggests accumulation of mobile elements into complex regions at or near junctions with tandemly repetitive DNA. Below we present a general model to account for this observation.

Discussion

Although the general characteristics of "satellite" sequences are well documented, the mechanisms governing their origin and maintenance still remain elusive. One of the reasons for not being able to resolve these fundamental processes is the inherent difficulties involved in studying long tandem arrays of repeated sequences. Analyses of junction sites at the ends of repetitive arrays provide an opportunity to characterize repetitive sequences with a unique level of resolution. Several studies of junctions between different tandemly repeated se-

quences reveal that these junctions are generally complex and not defined by a specific site of transition (Maresca and Singer 1983; Lohe and Brutlag 1987; Reed et al. 1994). This complexity arises because the different repeated sequences interdigitate throughout a junction region, suggesting that multiple recombination events have occurred between established repetitive arrays. Single junctions between repetitive and nonrepetitive sequences provide a better means for examining the properties underlying the evolution of tandem arrays, such as the studies of arrays of tandem repeats within coding regions (Eckert and Green 1986; Hogan et al. 1995), although it can be difficult to identify where the array begins. When junctions are not localized to a specific site, it is difficult to determine the potential role these regions may have played in the evolution of the repetitive array. The junction characterized in this study involves the tandem repeat PSR2 adjacent to a truncated end of the retroelement *NATE.* One benefit of this particular organization was the ability to resolve the transition site between *NATE* P5 and PSR2 by comparing the junction region to a full-length copy of *NATE* and a reference PSR2 repeat. This comparison revealed an extremely abrupt transition from the *NATE* sequence to the beginning of the PSR2 array, where only short regions of similarity were identified around the junction.

Origin of the Junction Site

By specifically identifying the junction site, this study was able to distinguish between sequences contained in the array and sequences flanking the array. Thus, the sequence of the full-length retroelement *NATE* P16 represents the ancestral state of *NATE* P5 prior to the origin of the junction. This study addresses several mechanisms that could have led to the origin of this junction between *NATE* P5 and PSR2. (1) The PSR2 sequence may have originated from the underlying *NATE* sequence. (2) During array amplification, tandemly repeated sequences may expand into nonrepetitive sequences that are flanking arrays. Therefore, the PSR2 array may be expanding into the sequence of *NATE* P5. (3) *NATE* P5 may have inserted into the PSR2 array or (4) alternatively the PSR2 array may have inserted into *NATE* P5. (5) The two sequences may have become juxtaposed following a structural rearrangement of the chromosome.

Several proposed mechanisms of tandem repeat origin make specific predictions about the relationship between the repeated sequence within an array and the sequence from which the array originated. Similarity between PSR2 and *NATE* is expected if the PSR2 repeat originated from this region of *NATE* by duplication (Krüger and Vogel 1975; Ohta 1980) or replication slippage (Levinson and Gutman 1987; Walsh 1987; Stephan 1989). Upon comparison of these sequences, no extensive similarity was detected between the PSR2 repeat and the

region of *NATE* underlying the junction (except for a short 11-bp region). Therefore, the PSR2 repeat was unlikely to have been generated from this region of *NATE* by sequence duplication or an aberrant replication error. Under the model of successive unequal exchanges causing the origin of a repetitive array, there is an unpredictable relationship between the final repeated sequence and the progenitor sequence (Smith 1976; Stephan 1989). Even though the relationship is not direct, a transition from the repeat sequence to the flanking sequence is expected, and this transition was not observed. Furthermore, instability of the junction site is expected if the array originates by successive unequal exchange, because chance unequal exchange events are expected to occur between repeats in the array and sequence flanking the array (Smith 1976). Inconsistent with this expectation is the apparent stability of the junction site among the five independent natural isolates of the PSR chromosome. Since this junction does not represent the expected structure following the primary origin of the PSR2 array by either duplication, slippage replication, or unequal exchange, another mechanism is necessary to explain the secondary formation of the junction following the establishment of the PSR2 repeat.

Secondary processes may have led to formation of this junction after the origin of the PSR2 array on this chromosome. One possibility is that *NATE* P5 inserted into the PSR2 array in its current form; however, two observations are inconsistent with this hypothesis. Since *NATE* P5 is truncated at one end, the long terminal repeat is absent from this end of the element making the current form of the element incapable of insertion (Varmus and Brown 1992). Upon initial insertion, *NATE* P5 must have contained long terminal repeats at both ends of the element and the truncation and formation of the current junction with PSR2 must have followed the element's insertion. Additionally, the PSR2 repeat is located only at the truncated end of *NATE* P5. Had this element inserted into a PSR2 array, the presence of the repeat is expected in the flanking regions at both ends of the element. However, it is possible that the PSR2 repeat was once located at both ends of the element, and the sequence has subsequently been lost on the intact end due to element accretion (see model below).

The immediate transition from the *NATE* P5 sequence to the PSR2 array is curiously similar to an insertion, and an alternative mode of insertion is that the PSR2 array inserted into this region of *NATE* P5. Although mobility of tandemly repeated sequences has never been demonstrated, the rolling-circle-replication model (Walsh 1987) provides a mechanism whereby this process could occur. The presence of extrachromosomal ''plasmids'' containing tandemly repeated sequences does provide empirical support that template for rolling circle replication is produced by intrastrand exchange within repetitive arrays (Kiyama et al. 1986, 1987; Okumura et al. 1987). How-

ever, replication of these plasmids and insertion of a linear array of repeats has not been demonstrated. The structure of the junction in this study is consistent with an insertion of the PSR2 array into *NATE* P5, but attempts at verifying this hypothesis by identifying the missing piece of *NATE* P5 at the other end of the PSR2 array were unsuccessful. Although a PSR2 insertion was not ruled out by the failure to detect the expected structure at the other end of the PSR2 array, it is unlikely that this structure is currently present on the PSR chromosome.

The absence of evidence supporting any of the previously mentioned mechanisms, and the presence of notable regions of similarity in this relatively short junction region, suggests that the junction was formed by nonhomologous recombination which joined an established PSR2 array with an internal region of *NATE* P5. When short regions of similarity are involved in nonhomologous recombination, resolution may result in products which are nonconservative (Sakagami et al. 1994), as appears in this junction between *NATE* P5 and PSR2. It is worth noting that one of the regions involved was in reverse orientation, therefore establishing a complex interaction among the DNA strands. This nonhomologous recombination would have truncated *NATE*. Since a large amount of this chromosome appears to be functionally inert (Beukeboom and Werren 1993; McAllister, Beukeboom, and Werren, in preparation), it is reasonable to assume that the rearrangement was not deleterious.

Maintenance and Evolution of Repetitive Arrays

This junction site is also useful for examining processes governing the maintenance and evolution of repetitive arrays. Unequal exchange is currently the favored mechanism to explain the observed sequence homogeneity among repeats in repetitive arrays (Krüger and Vogel 1975; Smith 1976; Stephan 1989; Charlesworth et al. 1994; Elder and Turner 1995). One prediction of the model is that repeats at the end of an array should exhibit the highest level of sequence divergence (Ohta 1980; Stephan 1989). This pattern has been documented in arrays of 200-bp tandem repeats within Balbiani ring genes in *Chironomus tentans* (Höög et al. 1988) and in alpha satellite arrays in the human genome (Wevrick et al. 1992; Cooper et al. 1993). This study indicates that the same phenomenon occurs in an array of PSR2 repeats on the PSR chromosome. The three most terminal repeats of this PSR2 array exhibited greater overall sequence divergence from a set of internal PSR2 repeats. Furthermore, these end repeats contained many unique nucleotide substitutions that have not been identified in other PSR2 repeats. Effects of the array terminus apparently act over a very short distance, because increased divergence was only observed in the first three repeats, whereas the fourth repeat was very similar to the internal

repeats. This observation is consistent with Stephan's (1989) simulation of repetitive sequence evolution where unequal exchange at the array terminus was suppressed by the presence of nonrepetitive sequences (i.e., presence of the nonhomologous *NATE* element), and this effect should operate over very short distances.

The observed pattern is consistent with unequal exchanges being inhibited at the array end. However, some unequal exchanges are apparent. Unequal exchange in these end repeats may be rare relative to exchange events within the array, and these exchange events may be localized (Ohta 1980). The data are consistent with both of these assumptions; only two unequal exchanges were evident and both of these apparently resulted from displacement of the array by a single repeat. However, why is there no evidence of gene conversion? How do the end repeats diverge in the presence of unequal exchange?

An alternative explanation that encompasses the observations is based on a revised model of unequal exchange proposed by Fletcher (1994). In his model, unequal exchange is initiated by a double-strand break in a repetitive array. To repair the break, the two strands search independently for homology and ultimately pair with different repeats. If the strands pair into a Holliday complex with the 5′ and 3′ ends in the correct orientation (facing each other), although with a number of repeats separating the ends, resolution of the complex causes expansion of the array. This junction region provides a special case for this process as illustrated by the inferred unequal exchange events in Fig. 4. When a double-strand break occurs near a junction site, homology searching and pairing of the strand containing the junction may be dominated by the flanking sequence, whereas the other strand contains repetitive sequence and can pair with any repeat in the array. Studies have demonstrated that recombination is influenced by long stretches of sequence identity (Waldman and Liskay 1987, 1988; Metzenberg et al. 1991), so when a break occurs near a junction the strand containing the array terminus and flanking sequence should preferentially anneal to its homologous site. Because the junction region and terminal repeats would always be repaired from identical sequence on the template DNA strand, the sequence of these repeats would be protected from gene conversion. Nucleotide substitutions occurring in the end repeats would be retained, even in the presence of recombination. Depending on where the break occurs, nucleotide substitutions present in the end repeats could shift into the array during this process, and the pattern exhibited by the last three repeats in this PSR2 array suggests that this has occurred.

The data on the windows in which recombination events have occurred also implicate the double-strand break model. Very short stretches of identity were observed in the regions where recombination was evident. The most recent inferred unequal exchange was localized

in a stretch of 26 bases of complete sequence identity, whereas the older exchange occurred in a window of 15 bases of complete identity between the junction with *NATE* and the distinct nucleotides that were shifted into the array by the exchange. Other studies have also revealed that unequal exchange in repetitive arrays occurs in very short tracts of identity (Cabot et al. 1993; Warburton et al. 1993; Reed et al. 1994). Because the double-strand break model allows the two broken ends to independently anneal at different locations on the template strand, the short window of identity where recombination events are localized may be misleading with regard to the sequence identity that was present between the broken ends and the template.

Findings of this study are consistent with unequal exchange being the mechanism maintaining homogeneity throughout the array. A previous report also documented unequal exchange on the PSR chromosome, although the exchanges were between different families of repetitive sequences (Reed et al. 1994). Because of the unique transmission pattern of PSR, unequal exchanges can only occur between sister-chromatids during mitotic (haploid) cell divisions, thus these studies provide support that unequal exchange can be limited to sister-chromatids. This is consistent with a sequence analysis of the *Responder* array in *Drosophila melanogaster* which indicated that unequal exchange was primarily an interchromatid process (Cabot et al. 1993).

### Accretion Model Spatial Structure of Tandem Arrays

We have demonstrated an association among tandemly repeated arrays and truncated dispersed repeats on the PSR chromosome, and here we propose a model that may provide a general explanation for this frequently observed association. It is recognized that mobile elements will accumulate within noncoding regions such as tandem DNA, because they are less likely to cause harmful mutations (Charlesworth et al 1994). However, there has been little discussion of how mobile elements are expected to be distributed spatially within tandem arrays. Here we present an ''accretion model'' that predicts the accumulation of mobile elements at the ends of arrays and into ''islands'' within arrays. The model is based on the prediction that, due to the stochastic processes of duplication and deletion from unequal exchanges, the ultimate fate of any particular array is loss (Charlesworth et al. 1986, Stephan 1987). This drift in array size has implications for macrostructural features of tandem arrays and associated DNAs. In particular, we propose that mobile elements inserting into tandem arrays will subsequently ''migrate'' to the ends of arrays, and into islands within arrays, due to the turnover of tandem repeats.

Consider an insertion of a mobile element into a tandem array (see Fig. 6). Due to the process of unequal
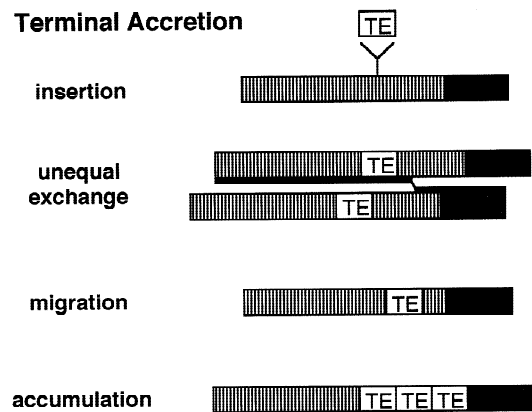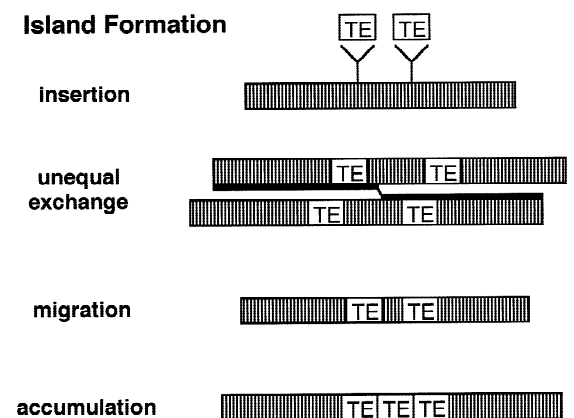
Fig. 6. Large-scale effects of mobile element insertion and unequal exchange on the structure of tandem arrays. Transposable elements (TE) insert into a large tandem array. A single product of unequal exchange is indicated by the *heavy line,* and this causes migration of the transposable element. Ultimately, accumulation of mobile elements occurs. **A** Terminal accretion of mobile elements at the end of an array. **B** Island formation by mobile element accumulation within repetitive arrays.

exchange, the portion of tandem array located to one side (or the other) of the element will eventually be lost by chance drift in repeat numbers, causing migration of the element to the end or the array (Fig. 6A). Following the insertion of more elements and repetition of the process, unequal exchange will ultimately cause the accumulation of many mobile elements at the ends of the tandem array. Similarly, if two elements are inserted into a tandem array (Fig. 6B), there would be an intervening array between them. Random loss of this intervening array would move the elements together where they would accumulate into an ''island'' within the tandem array by this accretion process. Accretion of additional elements and islands will lead to the further growth of non-array islands composed of dead, dying, and functional mobile elements.

The accretion model, therefore, predicts both islands of mobile elements within tandem arrays and large num-

bers of mobile elements at the ends of tandem arrays. Features in the cloned regions of the PSR chromosome are consistent with this prediction. In all regions containing the terminal region of a tandem array, many different middle repetitive sequences are present flanking these arrays. Retroelements should be more susceptible to accretion, because once inserted into a tandem array, they should rarely excise (Nuzhdin and Mackay 1994). Several recent studies have characterized complex repetitive DNAs involving retroelements (Wevrick et al. 1992; Hochstenback et al. 1994; Nurminsky et al. 1994; Le et al. 1995). Consistent with the accretion model, Wevrick et al. (1992) describe the presence of retroelements flanking alpha satellite arrays, and Le et al. (1995) found that *Drosophila* heterochromatin is composed of tandemly repetitive DNA with alternating regions of complex DNA. Analysis of one of the ''islands'' of complex DNA revealed presence of a retroposon.

Local accumulation of different mobile elements by accretion, coupled with truncations (as observed with copies of *NATE* on PSR), could result in the juxtaposition of different coding sequences. This may have implications for the evolution of novel genetic sequences, because it is conceivable that such accretions will be breeding grounds for new mobile elements, viruses, or novel coding sequences.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1992) Current protocols in molecular biology, Suppl 21. John Wiley & Sons, New York, pp. 7.3.7

Beukeboom LW, Werren JH (1993) Deletion analysis of the selfish *B* chromosome, *Paternal Sex Ratio* (PSR), in the parasitic wasp *Nasonia vitripennis.* Genetics 133:637–648

Bigot Y, Hamelin M-H, Periquet G (1990) Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). Mol Biol Evol 7:351–364

Cabot EL, Beckenbach AT (1989) Simultaneous editing of multiple nucleic acid and protein sequences using ESEE. Comp Appl Biosci 5:233–234

Cabot EL, Doshi P, Wu M-L, Wu C-I (1993) Population genetics of tandem repeats in centromeric heterochromatin: unequal crossing over and chromosomal divergence at the *Responder* locus of *Drosophila melanogaster.* Genetics 135:477–487

Chalker AF, Okeley EA, Davison A, Leach DRF (1993) The effects of central asymmetry on the propagation of palindromic DNA in bacteriophage λ are consistent with cruciform extrusion in vivo. Genetics 133:143–148

Charlesworth B, Langley CH, Stephan W (1986) The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics 112:947–962

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

Cooper KF, Fisher RB, Tyler-Smith C (1993) Structure of the sequences adjacent to the centromeric alphoid satellite DNA array on the human Y chromosome. J Mol Biol 230:787–799

Doolittle WF, Sapienza C (1980) Selfish genes, the phenotype paradigm and genome evolution. Nature 284:601–603

Eckert RL, Green H (1986) Structure and evolution of the human involucrin gene. Cell 46:511–523

Eickbush DG, Eickbush TH, Werren JH (1992) Molecular characterization of repetitive DNA sequences from a B chromosome. Chromosoma 101:575–583

Elder JF Jr, Turner BJ (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. Q Rev Biol 70:297–320

Fletcher HL (1994) Possible loss of length conservation and reciprocity during recombination or conversion in tandem arrays. Genetics 138:511–518

Hochstenback R, Harhangi H, Schouren K, Hennig W (1994) Degenerating *Gypsy* retrotransposons in a male fertility gene on *Y* chromosome of *Drosophila hydei.* J Mol Evol 39:452–465

Hogan NC, Slot F, Traverse KL, Garbe JC, Bendena WG, Pardue M-L (1995) Stability of tandem repeats in the *Drosophila melanogaster Hsr-omega* nuclear RNA. Genetics 139:1611–1621

Höög C, Daneholt B, Wieslander L (1988) Terminal repeats in long repeat arrays are likely to reflect the early evolution of Balbiani ring genes. J Mol Biol 200:655–664

Kiyama R, Hideki M, Oishi M (1986) A repetitive DNA family (*Sau*3A family) in human chromosomes: extrachromosomal DNA and DNA polymorphism. Proc Natl Acad Sci USA 83:4665–4669

Kiyama R, Okumura K, Matsui H, Bruns GAP, Kanda N, Oishi M (1987) Nature of recombination involved in excision and rearrangement of human repetitive DNA. J Mol Biol 198:589–598

Krüger J, Vogel F (1975) Population genetics of unequal crossing over. J Mol Evol 4:201–247

Le MH, Duricka D, Karpen GH (1995) Islands of complex DNA are widespread in Drosophila centric heterochromatin. Genetics 141:283–303

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol Biol Evol 4:203–221

Lohe AR, Brutlag DL (1987) Adjacent satellite DNA segments in *Drosophila,* structure of junctions. J Mol Biol 194:171–179

Maizel JV, Lenk RP (1981) Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc Natl Acad Sci USA 78:7665–7669

Maresca A, Singer MF (1983) Deca-satellite: a highly polymorphic satellite that joins α-satellite in the African green monkey genome. Mol Biol 164:493–511

McAllister BF (1995) Isolation and characterization of a retroelement from a B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis.* Insect Mol Biol 4:253–262

McAllister BF, Werren JH (1997) Hybrid origin of a B chromosome (PSR) in the parasitic wasp *Nasonia vitripennis.* Chromosoma 106:243–253

Metzenberg AB, Wurzer G, Huisman THJ, Smithies O (1991) Homology requirements for unequal crossing over in humans. Genetics 128:143–161

Nur U, Werren JH, Eickbush DG, Burke WD, Eickbush TH (1988) A ''selfish'' B chromosome that enhances its transmission by eliminating the paternal genome. Science 240:512–514

Nurminsky DI, Shevelyov YY, Nuzhdin SV, Gvozdev VA (1994) Structure, molecular evolution and maintenance of copy number of extended repeated structures in the *X*-heterochromatin of *Drosophila melanogaster.* Chromosoma 103:277–285

Nuzhdin SV, Mackay TFC (1994) Direct determination of retrotransposon transposition rates in *Drosophila melanogaster.* Genet Res 63:139–144

Ohta T (1980) Evolution and variation of multigene families. Springer-Verlag, New York

Okumura K, Kiyama R, Oishi M (1987) Sequence analyses of extra-chromosomal *Sau*3A and related family DNA: analysis of recombination in the excision event. Nucleic Acids Res 15:7477–7489

Orgel LE, Crick FHC (1980) Selfish DNA: the ultimate parasite. Nature 284:604–607

Reed KM, Beukeboom LW, Eickbush DG, Werren JH (1994) Junctions between repetitive DNAs on the PSR chromosome of *Nasonia vitripennis:* association of palindromes with recombination. J Mol Evol 38:352–362

Sakagami K, Tokinaga Y, Yoshikura H, Kobayashi I (1994) Homology-associated nonhomologous recombination in mammalian gene targeting. Proc Natl Acad Sci USA 91:8527–8531

Smith GP (1976) Evolution of repeated DNA sequences by unequal crossover. Science 191:528–535

Stephan W (1987) Quantitative variation and chromosomal location of satellite DNAs. Genet Res 50:41–52

Stephan W (1989) Tandem-repetitive noncoding DNA: forms and forces. Mol Biol Evol 6:198–212

Varmus H, Brown P (1989) Retroviruses. In: Berg DE, Howe MM (eds) Mobile DNA. Am Soc Microbiol, Washington, DC, pp. 53–108

Waldman AS, Liskay RM (1987) Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc Natl Acad Sci USA 84:5340–5344

Waldman AS, Liskay RM (1988) Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. Mol Cell Biol 8:5350–5357

Walsh JB (1987) Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115:553–567

Warburton PE, Waye JS, Willard HF (1993) Nonrandom localization of recombination events in human alpha satellite repeat unit variants: Implications for higher-order structural characteristics within centromeric heterochromatin. Mol Cell Biol 13:6520–6529

Werren JH, Nur U, Eickbush DG (1987) An extrachromosomal factor causing loss of paternal chromosomes. Nature 327:75–76

Wevrick R, Willard VP, Willard HF (1992) Structure of DNA near long terminal arrays of alpha satellite DNA at the centromere of human chromosome 7. Genomics 14:912–923

Willard, HF (1989) The genomics of long tandem arrays of satellite DNA in the human genome. Genome 31:737–744