Extremely complex repeat shuffling during germline mutation at human minisatellite B6.7

Keiji Tamaki, Celia A. May, Yuri E. Dubrova and Alec J. Jeffreys

Department of Genetics, University of Leicester, Adrian Building, University Road, Leicester LE1 7RH, UK

Received December 21, 1998; Revised and Accepted February 11, 1999 DDBJ/EMBL/GenBank accession no. AF081787

Human minisatellite B6.7 is a highly variable locus showing extensive heterozygosity with alleles ranging from six to >500 repeat units. Paternal and maternal mutation rates to new length alleles were estimated from pedigrees at 7.0 and 3.9% per gamete, respectively, indicating that B6.7 is one of the most unstable minisatellites isolated to date. Mutation at this locus was also analysed by small pool PCR of sperm and blood DNA. Male germline instability varied from <0.8 to 14% per allele and increased with tandem array size. In contrast, the frequency of mutants in somatic (blood) DNA was far lower (<0.5%), consistent with a meiotic origin of germline mutants. Sperm mutations were further characterized by minisatellite variant repeat mapping using four major polymorphic sites within the B6.7 repeats. This highly informative system revealed a wide variety of changes in allele structure, including simple inter-allelic duplications and deletions and more complicated inter- and intra-allelic transfers of repeat blocks, as seen at other human minisatellites. The main mode of sperm mutation, however, resulted in extremely complex allele reorganization with evidence of inter-allelic transfer plus the generation of novel repeats by rearrangement at the sub-repeat level, suggesting that recombinational instability at B6.7 is a complex multistep process.

INTRODUCTION

Minisatellites are a class of tandem repetitive DNA with 6 to >100 bp repeat units arranged in arrays ranging from 0.5 to 30 kb long. The most variable human minisatellites display a high frequency of spontaneous germline mutations altering repeat unit copy number and provide highly informative systems for analysing processes of tandem repeat turnover. De novo mutation at minisatellites can be detected by pedigree analysis (1) and much more efficiently by single molecule PCR of germinal (sperm) DNA, an approach which also enables the mutational behaviour of individual alleles to be explored (2). Analyses at human minisatellites MS32, MS205 and CEB1 have revealed significant variation between alleles in sperm mutation rate (2–4). At both MS32 and MS205, mutation rates appear to be independent of array length but modulated in cis; at MS32, a nucleotide transversion has been identified in the DNA flanking the minisatellite which is directly associated with a profound suppression of sperm instability, suggesting a role of flanking DNA in modulating repeat turnover (5). At CEB1, sperm instability can vary by three orders of magnitude between alleles, with the main factor influencing sperm mutation rate being array length (4).

Minisatellite mutation processes can be investigated by analysing the internal structure of new mutant alleles by minisatellite variant mapping by PCR (MVR-PCR) (6). Complex rearrangements, both intra-allelic duplications and polarized inter-allelic transfers of repeats, account for the vast majority of germline expansions at the unstable GC-rich minisatellites studied to date (2–4). At MS32, most mutants are derived from inter-allelic transfer and the insertion site of the repeat block is clustered within the first few repeats of the tandem array. These gene conversion-like rearrangements occur only in the germline, most likely at meiosis (7), and appear to be driven by a highly localized recombination hotspot adjacent to the repeat array that can occasionally yield true meiotic crossovers in and near the array (8,9). These findings raise the possibility that minisatellite instability in general may be a by-product of meiotic recombination in repeat DNA. Similar polarized germline mutation events (both intra- and inter-allelic) are observed at MS205, suggesting that these two minisatellites may share a common mutation/recombination pathway (3). Inter-allelic transfers at CEB1 similarly occur at a fairly constant rate irrespective of array length and show mild polarity. In contrast, intra-allelic rearrangements in CEB1 tend to cluster within sequence-homogeneous regions of the repeat array and their frequency increases with increasing array size (4).

Human minisatellite B6.7 was first discovered in 1992 by Kimpston et al. (10). The minisatellite consists of 34 bp repeats, is 56% GC-rich and has HinfI alleles ranging from 1 to 15 kb long with a reported allele length heterozygosity of 88% (10). Pedigree analysis indicated a mutation rate of several per cent (11,12), with a reported allele length heterozygosity of 88% (10). Pedigree analysis indicated a mutation rate of several per cent (11,12), with a more efficient approach enabling the mutational behaviour of individual alleles to be explored (2). Analyses at human minisatellites MS32, MS205 and CEB1 have revealed significant variation between alleles in sperm mutation rate (2–4). At both MS32 and MS205, mutation rates appear to be independent of array length but modulated in cis; at MS32, a nucleotide transversion has been identified in the DNA flanking the minisatellite which is directly associated with a profound suppression of sperm instability, suggesting a role of flanking DNA in modulating repeat turnover (5). At CEB1, sperm instability can vary by three orders of magnitude between alleles, with the main factor influencing sperm mutation rate being array length (4).

Received December 21, 1998; Revised and Accepted February 11, 1999 DDBJ/EMBL/GenBank accession no. AF081787

*To whom correspondence should be addressed. Tel: +44 116 252 3413; Fax: +44 116 252 3378; Email: kt12@le.ac.uk
RESULTS AND DISCUSSION

Sequence characterization around B6.7

The original B6.7 clone had limited flanking DNA and was therefore used to screen a human cosmids library. Clone cB6.7-1 was isolated containing a 36 kb insert and used to establish the sequence of the 75 repeat cloned allele plus 3 kb of flanking DNA. The flanking DNA contained a second minisatellite with 38–40 bp repeats plus an Alu element embedded in a mammalian transposon-like element long terminal repeat (MLT) element and a partial copy of an L1 (LINE) element (Fig. 1). This clustering of unstable minisatellites with other tandem repeats and dispersed repeats has been seen at other unstable minisatellite loci (13,14).

Size distribution of B6.7 alleles

We have a large collection of MboI-digested sperm DNA samples prepared for mutation analysis at minisatellite MS32 (2). Digestion with MboI releases the intact B6.7 minisatellite together with 846 bp of flanking DNA, allowing the size distribution of alleles to be determined from these digests by Southern blot analysis with the B6.7 probe. Apparent homozygotes were further analysed by PCR (see below), revealing that more than 65% of alleles are <100 repeats and only a small proportion are >200 repeats (Caucasian, >0.5% (upper 95% CI: 3.1–10.2%). Over-amplification of the smaller allele prevented analysis of limiting dilutions of genomic DNA. An example of SP-PCR analysis from an individual with alleles of 145 and 44 repeats is shown in Figure 2. Thirty-five mutant molecules with aberrant repeat copy number derived from the smaller allele were detected in sperm DNA, giving a germline mutation rate of 6.2% (95% CI: 3.1–10.2%).

Allele-specific sperm mutation rates were estimated for 34 alleles from 22 men (three Asian, two African and 17 Caucasian) by SP-PCR. The origin of each mutant was deduced by assuming that it was derived from the progenitor allele closer in size (within ±20 repeats). Structural analysis of mutant alleles detected in two men confirmed that almost all mutants (96%) had been assigned to the correct progenitor (see below). For 10 men, only the shorter allele was analysed since mutants could not be reliably scored for the larger allele (>90 repeats). For six alleles ranging from 58 to 93 repeats, changes of ±1 or 2 repeats could not be reliably scored. A combined mutation spectrum was therefore deduced from the remaining 28 alleles (Fig. 3). All alleles, whether small or large, showed a very similar size range of mutants involving small length changes and with a strong (72%) bias towards gains rather than losses of repeats. Similar gain biases have also been seen at

Mutation rates in pedigrees

CEPH pedigrees were screened by Southern blot hybridization for B6.7 length mutations. To achieve maximal resolution of alleles by minimizing the amount of flanking DNA, each DNA was digested with AluI. Thirty length-changed mutants were detected in 274 children. Eighteen of these mutants were paternal in origin and 10 maternal; the remaining two mutants were in families where allele sharing by the mother and father prevented the parental origin of mutation from being determined. The paternal and maternal mutation rates are 7.0 and 3.9% per gamete, respectively, similar to initially reported values (11), and confirming that B6.7 is one of the most unstable minisatellites isolated to date with an unusually high mutation rate, in particular in the female germline.

Most minisatellite mutations, including those at B6.7 (see below), involve the gain or loss of relatively small numbers of repeats (2–4). The direction of B6.7 mutation was, therefore, deduced by comparing mutant allele sizes with the sizes of progenitors. For paternal mutants, nine were gains and four losses, with five being intermediate in size between the two paternal alleles and therefore of uncertain origin. For maternal mutation, there were no gains, seven losses and three of uncertain origin. These gain/loss distributions for paternal and maternal mutation are significantly different (Fisher’s exact test, P < 0.01), suggesting differences in the mutation process between the male and female germlines.

Detecting mutants by SP-PCR

SP-PCR was used to investigate repeat instability in sperm and somatic (blood) DNA. Primers 67A and 67B flanking the minisatellite (Fig. 1) were used to amplify B6.7 molecules from multiple aliquots of MboI-digested genomic DNA containing 30–60 molecules of each progenitor allele. The number of amplifiable molecules was estimated by Poisson analysis of limiting dilutions of genomic DNA. An example of SP-PCR analysis from an individual with alleles of 145 and 44 repeats is shown in Figure 2. Thirty-five mutant molecules with aberrant repeat copy number derived from the smaller allele were detected in sperm DNA, giving a germline mutation rate of 6.2% (95% CI: 3.1–10.2%). Over-amplification of the smaller allele prevented the reliable detection of mutants at the larger allele. In contrast, analysis of 650 amplifiable progenitor molecules from blood DNA revealed no mutants, indicating a mutant frequency of <0.5% (upper 95% CI) per progenitor molecule, at least 13-fold lower than in sperm. This establishes that B6.7 mutation is largely restricted to the germline and confirms that most or all mutants recovered from sperm DNA are authentic and not PCR artefacts. Germline specificity of mutation has also been found for several other unstable human minisatellites (2–4).

Allele-specific variation in sperm mutation rate

Allele-specific sperm mutation rates were estimated for 34 alleles from 22 men (three Asian, two African and 17 Caucasian) by SP-PCR. The origin of each mutant was deduced by assuming that it was derived from the progenitor allele closer in size (within ±20 repeats). Structural analysis of mutant alleles detected in two men confirmed that almost all mutants (96%) had been assigned to the correct progenitor (see below). For 10 men, only the shorter allele was analysed since mutants could not be reliably scored for the larger allele (>90 repeats). For six alleles ranging from 58 to 93 repeats, changes of ±1 or 2 repeats could not be reliably scored. A combined mutation spectrum was therefore deduced from the remaining 28 alleles (Fig. 3). All alleles, whether small or large, showed a very similar size range of mutants involving small length changes and with a strong (72%) bias towards gains rather than losses of repeats. Similar gain biases have also been seen at
Figure 2. SP-PCR analysis of B6.7 mutation in sperm and blood DNA from a single individual. Each PCR reaction contained 26 amplifiable molecules of the lower allele (44 repeats in size).

MS32, MS205, MS31A and CEB1 (2–4). This constancy of mutation spectrum over different alleles allowed mutation rates to be corrected for unresolved small events at the remaining six alleles. The average mutation rate over all 34 alleles was 4.6% (95% CI: 4.1–5.2%). This rate is not significantly different from the paternal mutation rate of 7% (95% CI: 4–10%) established in pedigrees. The slight apparent reduction of rate may be due to the sperm samples being biased towards smaller and more readily analysed alleles that tend to show lower mutation rates (see below).

B6.7 alleles show considerable variation in instability, with observed sperm mutation rates varying from <0.8 to 14.7%. Rates are strongly correlated with allele size (Fig. 4) and increase steadily with array size up to ~50 repeats, beyond which they appear to reach a plateau. Below 20 repeats, the mutation rate decreases dramatically to <0.5% for the shortest alleles with <10 repeats. The relationship between array length and instability is very similar to that previously established for minisatellite CEB1 (Fig. 4b) (4). Array length is not the only variable influencing B6.7 instability. Thus, one African individual with two medium size alleles (73 and 53 repeats corresponding to R1U and R1L, respectively, in Fig. 4a) showed unusually low sperm mutation rates at both alleles (<1.5 and 0.8%, respectively; Fig. 4a). It is possible that this man happens to carry two very stable alleles with mutation suppressed in cis on each allele, analogous to minisatellite MS32 alleles that have been shown to be stabilized in cis (5). Alternatively, there may be a more general suppression of repeat instability in trans, as also suggested by the low CEB1 mutation rate seen in this same man (data not shown).

Developing MVR-PCR at B6.7

Further analysis of mutation processes at B6.7 required the development of MVR-PCR systems for characterizing allele structure before and after mutation. Thirty-nine B6.7 repeats derived from the cosmid clone and from four different small

Figure 3. Distribution of sperm mutant allele lengths. Twenty-eight alleles for which all mutants could be reliably scored were divided into two groups by size and the relative frequency of each size class of mutant determined. Both groups have a strong bias towards gains ($\chi^2$, df = 1, 19.40 and 91.38 for small and large alleles, respectively, $P << 0.0001$).
Figure 4. Relationship between mutation rate and allele length. (a) Mutation rates were estimated for 34 different B6.7 alleles by SP-PCR. Two alleles from the same individual (marked R1U and R1L) showed significantly low mutation rates. Quadratic fit for the observed mutation rate (solid line) is $m = -2.00 + 0.2917 \times \text{size} - 0.0024 \times \text{size}^2$, $r^2 = 0.465$, $F(2,31) = 13.48$, $P = 6.14 \times 10^{-5}$. For six alleles containing 58–93 repeats (open circles), gain or loss of one or two repeats could not be reliably scored and mutation frequencies were therefore corrected using the combined mutation spectrum from 16 large alleles (Fig. 3). The quadratic fit for the corrected mutation rate (dashed line) is $m = -1.94 + 0.2806 \times \text{size} - 0.0021 \times \text{size}^2$, $r^2 = 0.493$, $F(2,31) = 15.05$, $P = 2.71 \times 10^{-5}$. (b) Relationship between sperm mutation rate and allele length at B6.7 and CEB1 (4). Filled triangles indicate the mean mutation rates of the number of the alleles ($n$) within a given size class and the associated 95% confidence intervals (CI) are indicated by vertical lines.

Alleles amplified from genomic DNA were therefore sequenced. Three base substitutional polymorphic sites (positions 7, 11 and 18 in the repeat) and one polymorphic base deletion at position 5 were identified (Fig. 5a). There was no obvious association between polymorphic sites, except for T at position 7 occurring whenever there is a deletion at position 5. Two less common base substitutional variants were also noted at positions 14 and 21 (see sequence in Table 1 legend). Forward four-state MVR-PCR mapping and reverse three-state mapping strategies were developed to detect all four common polymorphic sites. The sequences of primers used are shown in Figure 5b. Position 11 and 18 polymorphisms are detected simultaneously by four-state forward MVR-PCR mapping. Position 5 and 7 polymorphisms are detected by reverse MVR-PCR; the deletion at position 5 is detected by primer 67TAG-NR and primers 67TAG-CR and 67TAG-TR detect the position 7 polymorphism provided that there is no deletion at position 5. The status at position 7 is unknown when there is a deletion at position 5, although sequence data suggest that it is likely to be T.
Figure 5. Repeat sequence variation at B6.7. (a) Sequence of the B6.7 repeat unit. The sequence corresponding to 1.5 34 bp repeat units shows the positions of four polymorphic sites and the location of forward and reverse MVR-specific primers (arrows). (b) PCR primer sequences used in the present study. Positions within tagged primers which are specific to minisatellite variant repeats are shown in bold.

Figure 6 shows MVR-PCR products derived from the cloned B6.7 minisatellite. MVR codes are read from the bottom in forward mapping and the top in reverse mapping. Forward and reverse MVR codes were then combined to deduce the sequence status of each repeat unit. The resulting code completely matched that predicted from the sequence data. However, combining forward and reverse mapping information is only possible if an allele is mapped completely in both directions; in practice this requires the allele to be >100 repeats.

The two MVR-PCR systems can in theory provide 20 different MVR codes if null repeats are included which contain additional variants that prevent MVR-specific primers from annealing (18). In practice, 18 different states were detected in eight different B6.7 alleles mapped by MVR-PCR, of which five were common (Table 1). These variant repeats were heavily interspersed along B6.7 alleles (Fig. 7), providing a rich source of internal structural variation for mutation analysis.

MVR analysis of sperm mutants

Two men (Y and Z) were selected for sperm mutation analysis. Sperm mutants derived from the two alleles (Y/y, Z/z) in each man were recovered by size fractionation of sperm DNA followed by SP-PCR (see Materials and Methods) and analysed by MVR-PCR (Fig. 7). Sixty-nine mutants were characterized in total. The 20-state MVR mapping system was sufficiently informative to allow the precise mapping of deletion breakpoints and the location of repeat insertions.

Table 1. Variant repeats at minisatellite B6.7

| MVR code | Position | No. of repeats (%) |
|----------|----------|-------------------|
| a        | T C C G   | 43 (18.5)         |
| b        | T T C G   | 41 (17.6)         |
| c        | N C G     | 2 (0.9)           |
| d        | T C A G   | 2 (0.9)           |
| e        | T T A G   | 50 (21.5)         |
| f        | N A G     | 35 (15.0)         |
| g        | T C C A   | 8 (3.4)           |
| h        | T T C A   | 3 (1.3)           |
| i        | N C A     | 1 (0.4)           |
| j        | T C A A   | 0 (0.0)           |
| k        | T T A A   | 8 (3.4)           |
| l        | N A A     | 14 (6.0)          |
| m        | T C u u   | 4 (1.7)           |
| n        | T T u u   | 6 (2.6)           |
| p        | N u u     | 4 (1.7)           |
| q        | u u C G   | 2 (0.9)           |
| r        | u u A G   | 5 (2.1)           |
| s        | u u C A   | 1 (0.4)           |
| t        | u u A A   | 0 (0.0)           |
| o        | u u u     | 4 (1.7)           |

Proportions of variant repeats determined by MVR mapping of 233 repeats from eight different alleles. Variant repeats are classified into 20 different states including the fully ‘null’ state (o) which fails to amplify with any of the five MVR primers. B6.7 repeat sequence: 5’-TCTC|TA Y|AGGMCAYGAGRGKGA-CAGTGAGGGGG-3’ (the four polymorphic positions are shown in bold).

Simple gain mutants. Forty-eight gain mutants were mapped (Fig. 7). Almost all mutants were different, as expected for products of meiotic repeat instability. The only exceptions were mutants y9 and y10, which show an intra-allelic duplication within a sequence-homogeneous region of allele y. It is unclear whether this represents germinal mosaicism for a premeiotic mutant or whether instead this region of allele y is prone to recurrent meiotic instability. There were three other examples of simple intra-allelic duplications (mutants Y3, Y10 and Z16). One mutant (y2) showed a recombinant repeat array commencing with one allele and terminating with repeats from the other allele, although the mutant structure is complex with seven anomalous repeats at the putative crossover breakpoint. It remains unclear whether this is a true meiotic crossover event of the type seen at minisatellite MS32 (8,9) or whether instead it has arisen by a more complex recombinational interaction between alleles Y and y. Several other mutants (e.g. z5, z4 and z5) had clearly arisen by a gene conversion-like transfer of blocks of repeats between alleles, either without (z5) or with (y5 and z4) target site duplication of repeats in the recipient allele. These conversions are similar to those previously seen at minisatellites MS31A, MS32, MS205 and CEB1 and suggest that a substantial component of repeat
instability at all five loci involves aberrant meiotic recombination (2–4).

A model for male germline repeat instability at CEB1 has been proposed previously (19). Two staggered single-strand nicks in the repeat array result in a double-strand break (DSB) with 3′ protruding single strands. The break is repaired by sister chromatid exchange or by pairing with the homologous allele. In the case of gain mutants, this produces two heteroduplex regions flanking the insertion. Differential repair of these heteroduplexes can result in three types of gain mutant: (i) simple insertions (type I); (ii) duplications of recipient motifs around the insertion (type II); and (iii) deletions of recipient motifs next to the insertion point (type III). Some B6.7 mutants can be explained by this model (e.g. type I, Z5; type II, Y5; type III, Y8). It is also possible that the heteroduplex regions can be repaired by alternating between the two allele templates (Fig. 8). For example, mutant y6 has an apparent inter-allelic transfer of repeats although there is an anomalous b repeat in the middle of the inserted block. This repeat can be explained as the result of patchwork heteroduplex repair using both the inserted strand (a repeat) and the recipient strand (e repeat) to create a recombinant b repeat (Fig. 8b).

Complex gain mutants. Surprisingly, the majority (65%) of B6.7 gain mutants have acquired a considerable number (1–17) of repeats for which there are no contiguous matches in either progenitor allele and whose origin is therefore unclear. Relatively complex events have also been seen at minisatellites MS32, MS31A and CEB1 (2,4), but not with the frequency and complexity that characterize B6.7 mutation. Like the conversion events, these B6.7 transfers can be accompanied by recipient allele duplication flanking (e.g. Y7) or near (e.g. Y4) the site of insertion or in loss of information from the recipient (e.g. Y11).

At least some of these complex mutants have arisen by a process that includes inter-allelic transfer of information. For example, mutant y1 contains an ac motif transferred in register from allele Y to allele y, followed by 16 anomalous repeats including b, q and r repeats found only in allele Y. These complex transfers are more evident in mutants derived from the small alleles which have a more restricted repertoire of repeat types. As a result, the origin of complex insertions into the larger alleles remains enigmatic.

At least 18 of the 48 gain mutants cannot be readily explained by the staggered nick mutation model for CEB1 (19). In many cases the length of the inserted block is too short to predict its origin, but in other cases it is impossible to generate the insertion even through very fine-scale patchwork heteroduplex repair. The recent observation in yeast that the template of gap repair consists of two unlinked overlapping sequences has led to the synthesis-dependent strand annealing (SDSA) model of recombination (20) being modified to incorporate reinvasion by the 3′ protruding end (21). Whilst a few mutants of complex repeat rearrangements at B6.7 may be explained by this model (e.g. Y5, Z11), there are a considerable number of mutant structures that still remain uninterpretable. One possible process represents an extension of SDSA: mutation is initiated by cleavage of the recipient allele which then undergoes multiple rounds of strand invasion and repair synthesis with either the same molecule, the sister chromatid or homologous allele until eventually the DSB is bridged by a patchwork of information derived from both alleles. Such stepwise repair could also create novel repeats not present in either progenitor, as seen for example in mutants y8 and Y11 (e repeats).

MVR-PCR does not reveal all sequence variation within B6.7 alleles. To characterize mutants further, both alleles in individual Z together with five mutants marked in bold in Figure 7 were completely sequenced (data not shown). The sequence information confirmed that the MVR data are authentic and that there can be additional internal sequence variation even within the same MVR repeat type (e.g. the sequence of the m repeat at position 5 is different from the other m repeats at positions 29 and 32 in allele Z). To some extent these sequence data helped to further analyze repeat rearrangement in mutants. For example, it revealed that z4 has an inter-allelic event with duplication of the m (or ma) repeat(s) of the recipient allele on both sides of the inserted block and that the inserted block itself appears to contain a deletion of 29 repeat units of the donor allele. However, sequencing did not clarify the origin of the more complex mutants.

Complex rearrangements, as well as simpler inter-allelic conversions and intra-allelic rearrangements, are sperm specific and do not occur at a significant frequency in the soma. It is likely that all three classes of mutation arise by alternative processing.

Figure 6. Forward and reverse MVR-PCR mapping of B6.7 allele structure. Aliquots of 12 pg of cosmid cB6.7-1 DNA were MVR mapped using PCR primers and conditions described in Materials and Methods. This allele contains 75 repeats.
Figure 7. Structures of sperm gain mutants determined by MVR-PCR. Two men (Y and Z), heterozygous for alleles Y/y and Z/z, respectively, were analysed. Repeats from the smaller allele are shown in red and those from the larger allele in green. Repeats of no obvious origin are indicated in blue. Intra-allelic duplications are indicated by double underlining of the 5′ duplicated unit and single underlining of the 3′ unit. Distinctive intra-allelic inserted blocks are underlined with dots.

of the same recombination initiation complex. To test this, the location of each class of event along alleles was analysed (Fig. 9a). Complex rearrangements and intra-allelic events showed very similar distributions, consistent with a common origin. There was no clear evidence for polarity as seen for mutation at minisatellites MS31A, MS32 and MS205 (2,3), although events in allele Y tended to be displaced towards the 3′-end (primer 67B-proximal) of the array. The relatively few definite examples
Figure 8. The possible involvement of patchwork heteroduplex repair in the generation of mutant y6. (a) A model for mutation based on staggered nicks (19). Nicks (arrowed) are introduced into one allele (i), resulting in a double-strand break with 3′ protruding single-strand overhangs (ii). One overhang invades the homologous allele in register (iii) and is extended (iv). The extended strand is extruded from the homologous allele and anneals with the other 3′ overhang, bridging the double-strand break (v). The remaining single-strand gap is filled in by repair synthesis (vi). The resulting molecule contains a region of heteroduplex (boxed) which is repaired using both strands to create the novel mutant structure (vii). (b) Patchwork repair of the heteroduplex region. Both strands of the heteroduplex and the corresponding region of the mutant are shown. Each repeat is represented by a box; the state of each of the three polymorphic positions assayed by MVR-PCR (positions 5, 7 and 11; Table 1) is shown within the box and the corresponding single-letter MVR code outside. The strand used for heteroduplex repair is shown as a solid line for positions that differ between the two strands and as a dashed line where information could have been derived from either allele. The anomalous b-type repeat therefore arises by repair switching to the strand from the recipient allele over a region at least 1 nt long but ≤73 nt long.

of inter-allelic conversion detected mainly on the smaller alleles y and z again showed no pronounced polarity, although there was some evidence for displacement towards the 5′-end of the array.

Deletion events. Deletions are less common than insertions and none were detected for allele z. Twenty-one sperm deletion mutants from the other three alleles were characterized (data not shown). Seventeen of these were simple, involving the loss of a single contiguous block of repeats. As for insertions, some repeat types at the deletion breakpoint could be interpreted as a hybrid repeat formed from the 5′-region of the first repeat and the 3′-region of the last repeat of the deleted repeat block. Two deletion mutants exhibited a single MVR change that was apparently unrelated to the deletion event; these could be
on 25 July 2018

Has been established at minisatellite MS32 (8,9), remains to be a multistep mechanism that can profoundly reshuffle the relationship between B6.7 repeat instability processes to other unstable VNTR loci, with strong evidence for Minisatellite B6.7 shows similarities in overall sperm mutation polarity. This is similar to deletion mutant distribution at CEB1 (19) and contrasts with deletion polarity seen at other minisatellites (2,3).

MATERIALS AND METHODS

Samples

Human DNA samples were prepared from sperm and blood as described previously (2).

Cosmid isolation and sequencing

Clones (3 × 10^5, three genome equivalents) from a human placental DNA cosmid library cloned in pWE15 (Clontech, Palo Alto, CA) were screened with the B6.7 repeat probe (P. Gill, Forensic Science Service, UK) using Hybond Nfp membranes. The human DNA insert from the single clone isolated was sonicated, shotgun cloned into pBluescript II KS+ (Stratagene, La Jolla, CA) and selected subclones sequenced using either single-stranded phagemid DNA or PCR-amplified double-stranded phagemid inserts according to the protocol supplied from Perkin Elmer (ABI PRISM Dye Terminator Cycle Sequencing Ready kit; Perkin Elmer, Foster City, CA).

SP-PCR and detection of mutants

SP-PCR was carried out as described previously (2) with minor modifications. Samples of 180–360 pg MboI-digested sperm or blood DNA were amplified in 7 µl reactions using the PCR buffer system described previously (22) plus 0.2 µM primer 67A, 0.2 µM primer 67B (Fig. 1) and 0.1 U/µl Taq polymerase. Reactions were cycled for 45 s at 96°C, 45 s at 68°C and 3 min at 70°C for 26 cycles on a GeneAmp PCR system 9600 Thermal Cycler (Perkin Elmer). Aliquots of each SP-PCR reaction (2 µl) were electrophoresed through a 35 cm 1% agarose gel (Seakem LE; FMC BioProducts, ME) in 0.5× TBE (44 mM Tris–borate, pH 8.3, 1 mM EDTA), blotted onto Hybond Nfp membranes. The human DNA insert from the single clone isolated was sonicated, shotgun cloned into pBluescriptII KS+ (Stratagene, La Jolla, CA) and selected subclones sequenced using either single-stranded phagemid DNA or PCR-amplified double-stranded phagemid inserts according to the protocol supplied from Perkin Elmer (ABI PRISM Dye Terminator Cycle Sequencing Ready kit; Perkin Elmer, Foster City, CA).

MVR-PCR of B6.7 alleles

The general strategy for MVR-PCR (6) uses a combination of a fixed primer in DNA flanking the minisatellite and repeat variant-specific primers. Detection of a given repeat variant is achieved in the first few cycles by the MVR-specific primer which is used at low concentration. This primer carries a 5′ extension or ‘TAG’ which serves to subsequently drive amplification as the flanking primer and the TAG sequence itself are present at high concentrations. Forward MVR-PCR was performed using MVR primers 67TAG-CG, 67TAG-CT, 67TAG-

Figure 9. Distribution of sperm mutation events along B6.7 alleles. (a) Relative location of gain events, with alleles scaled to the same length. Filled circles, inter-allelic conversions; hatched circles, complex events; open circles, intra-allelic rearrangements. For rearrangements involving duplications in the recipient allele, the location is given as the mid-point of the duplication. (b) Relative location of deletion mutants. The extent of each deletion is shown below as a horizontal line, with the centre point and progenitor allele indicated above. The bottom two mutants each showed the loss of two separate repeat blocks. Two mutants showing one block deletion plus a change of one MVR repeat type elsewhere in the array are marked with asterisks.

Interpreted as microconversions of either intra- or inter-allelic origin. Two mutants showed more complex changes involving the loss of two separate blocks. None of the deletion mutants appeared to have arisen through crossover events. The distribution of deletions along alleles (Fig. 9b) showed no evidence of polarity. This is similar to deletion mutant distribution at CEB1 (19) and contrasts with deletion polarity seen at other minisatel-lites (2,3).

Final remarks

Minisatellite B6.7 shows similarities in overall sperm mutation processes to other unstable VNTR loci, with strong evidence for repeat instability arising through aberrant processing events during meiotic recombination. The remarkable complexity of many B6.7 mutants suggests that processing of a recombination initiation complex within tandem repeat DNA is not simple and can involve a multistep mechanism that can profoundly reshuffle patterns of repeat units to create major rearrangements in a single mutation event. The relationship between B6.7 repeat instability and true meiotic crossover events in and near the minisatellite, as has been established at minisatellite MS32 (8,9), remains to be investigated. Unlike most other minisatellites, and especially CEB1 (16), B6.7 also shows substantial maternal instability, allowing for the first time the possibility of detailed analysis of mutation processes in the female germline.
TG and 67TAG-TT at final concentrations of 2, 2, 8 and 8 nM, respectively. The driver primers (67A, 67C and TAG) were used at 0.4 µM. PCR reactions (7 µl, containing 100 ng genomic DNA) were cycled for 45 s at 96°C, 1 min at 68°C and 4 min at 70°C for 18 cycles. In reverse MVR-PCR, the MVR-specific primers 67TAG-TR, 67TAG-TG and 67TAG-NR were used at concentrations of 2 nM, in conjunction with driver primers 67B and TAG at 0.4 µM, using the same cycling conditions as for forward MVR-PCR. PCR products were resolved by agarose gel electrophoresis and detected by Southern blot hybridization with 32P-labelled B6.7 probe.

Size fractionation of genomic DNA

Fractionation was carried out as described previously for MS32 with minor modification (7). Samples of 4 µg of sperm DNA were digested to completion with MboI and electrophoresed through a 20 cm 1% SeaKem HGT (FMC BioProducts) agarose gel in 0.5x TBE buffer at 40 V for 16 h. Molecular weight markers (1 kb DNA ladder; Life Technologies, Paisley, UK) were visualized by staining with ethidium bromide and used to locate the expected position of the progenitor allele(s). The genomic DNA lane was fractionated into gel slices ranging from 1.5 (close to the progenitor) to 3 mm thick (other regions), to maximize enrichment of mutants similar in size to the progenitor allele. Each gel slice was then crushed in 70 µl of water and the DNA released by three cycles of freezing and thawing. After centrifugation at 12,000 r.p.m. (9400 g) for 2 min, the supernatant was used as a source of input DNA for SP-PCR detection of mutant molecules. This fractionation simplified the recovery of mutants, in particular for long mutants in the presence of a short progenitor allele that would otherwise differentially overamplify during SP-PCR.

Isolation and MVR analysis of sperm mutants

For each pool of fractionated DNA containing mutant molecules, 0.5 µl of SP-PCR products was reamplified for five cycles and fractionated by electrophoresis through a 35 cm 1% SeaKem HGT agarose gel in 0.5x TBE buffer at 120 V for 14 h. Using molecular weight markers, gel slices spanning the predicted position of the mutant allele were excised, DNA released by freezing/thawing and reamplified by semi-nested PCR using primers 67B and 67C (0.2 µM) for 27 cycles as above. Mutant bands, now detectable by ethidium bromide staining after gel electrophoresis, were recovered from the gel slices and aliquots corresponding to ~40 pg DNA were analysed by MVR-PCR.

ACKNOWLEDGEMENTS

We are grateful to P. Gill for supplying the original B6.7 probe and to J. Blower, A.D. Nkomo, S.B. Kanoyangwa and volunteers for providing semen samples. We thank K. Lilley and S. Bayliss for oligonucleotide synthesis and assistance with automated sequencing. Ila Patel for helping with Southern blot analysis and Jérôme Buard, Philippe Bois and colleagues for helpful advice. K.T. is a Wellcome Trust Research Fellow. This work was supported in part by an International Research Scholars Award to A.J.J. from the Howard Hughes Medical Institute and in part by grants from the Wellcome Trust, Medical Research Council and Royal Society.

REFERENCES

1. Jeffreys, A.J., Royle, N.J., Wilson, V. and Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature, 338, 278–281.

2. 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.

3. May, C.A., Jeffreys, A.J. and Armour, J.A.L. (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. Nature Genet., 8, 162–170.

4. Jeffreys, A.J., MacLeod, A., Tamaki, K., Neil, D.L. and Monckton, D.G. (1991) Minisatellite repeat coding as a digital approach to DNA typing.

5. Buard, J., Boudet, A., Yardley, J., Dubrova, Y. and Jeffreys, A.J. (1998) Influences of array size and homogeneity on minisatellite mutation. EMBO J., 17, 3495–3502.

6. Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A. and Jeffreys, A.J. (1994) Minisatellite mutation rate variation in human minisatellites arising from meiotic recombination. EMBO J., 17, 4147–4157.

7. Jeffreys, A.J., Neil, D.L. and Neumann, R. (1998) Repeat instability at human minisatellites arising from meiotic recombination. EMBO J., 17, 4147–4157.

8. Jeffreys, A.J., Neumann, R. (1997) Somatic mutation processes at a human minisatellite. Hum. Mol. Genet., 6, 129–136.

9. Jeffreys, A.J., Neil, D.L. and Neumann, R. (1998) Repeat instability at human minisatellites arising from meiotic recombination. EMBO J., 17, 4147–4157.

10. Kimpton, C.P., Hopgood, R., Watson, S.K., Gill, P. and Sullivan, K. (1992) Mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA.

11. Olaisen, B., Bekkemoen, M., Hoff-Olsen, P. and Gill, P. (1993) Human minisatellite mutation and sex. In Pena, S.D.J., Chakraborty, R., Epplen, J.T. and Jeffreys, A.J. (eds), DNA Fingerprinting: State of the Science. Birkhauser Verlag, Basel, Switzerland, pp. 63–69.

12. Dubrova, Y.E., Nesterov, V.N., Krouchinsky, N.G., Ostapenko, V.A., Vergnaud, G., Giraudieu, P., Buard, J. and Jeffreys, A.J. (1997) Further evidence for elevated human minisatellite mutation rate in Belarus eight years after the Chernobyl accident. Mutat. Res., 381, 267–278.

13. Armour, J.A.L., Wong, Z., Wilson, V., Royle, N.J. and Jeffreys, A.J. (1989) Sequences flanking the repeat arrays of human minisatellites: association with tandem and dispersed repeat element. Nucleic Acids Res., 17, 4925–4935.

14. Murray, J., Buard, J., Neil, D.L., Yeramian, E., Tamaki, K., Hollies, C. and Jeffreys, A.J. (1999) Comparative sequence analysis of human minisatellites showing meiotic repeat instability. Genome Res., 9, 136–145.

15. Armour, J.A.L., Harris, P.C. and Jeffreys, A.J. (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. Hum. Mol. Genet., 2, 1317–1345.

16. Vergnaud, G., Mariat, D., Apiou, F., Aurias, A., Lathrop, M. and Lauthier, V. (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. Genomics, 11, 135–144.

17. Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A.J. (1987) Characterization of a panel of highly variable minisatellites cloned from human DNA. Ann. Hum. Genet., 51, 269–288.

18. Tamaki, K., Monckton, D.G., MacLeod, A., Neil, D.L., Allen, M. and Jeffreys, A.J. (1992) Minisatellite variant repeat (MVR) mapping: analysis of 'null' repeat units at D1S8. Hum. Mol. Genet., 1, 401–406 (Erratum: 1992; Nature, 354, 558).

19. Buard, J. and Vergnaud, G. (1994) Complex recombination events at the hypervariable minisatellite CEB1 (D2S80). EMBO J., 13, 3203–3210.

20. Nassif, N., Penney, J., Pal, S., Engels, W.R. and Gloor, G.B. (1994) Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. Mol. Cell. Biol., 14, 1613–1625.

21. Paques, F., Leung, W.-Y. and Haber, J.E. (1998) Expansions and contractions in a tandem repeat induced by double-strand break repair. Mol. Cell. Biol., 18, 2045–2054.

22. Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell, 60, 473–485.