



Minisatellites: Mutability and Genome Architecture

Gilles Vergnaud and France Denoeud

Genome Res. 2000 10: 899-907

Access the most recent version at doi:[10.1101/gr.10.7.899](https://doi.org/10.1101/gr.10.7.899)

References This article cites 51 articles, 13 of which can be accessed free at:
<http://genome.cshlp.org/content/10/7/899.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Minisatellites: Mutability and Genome Architecture

Gilles Vergnaud^{1,2,3} and France Denoeud¹

¹*Institut de Génétique et Microbiologie, Université Paris Sud 91405 Orsay, France;* ²*Centre d'Etudes du Bouchet, 91710 Vert le Petit, France*

Minisatellites have been found in association with important features of human genome biology such as gene regulation, chromosomal fragile sites, and imprinting. Our knowledge of minisatellite biology has greatly increased in the past 10 years owing to the identification and careful analysis of human hypermutable minisatellites, experimental models in yeast, and recent *in vitro* studies of minisatellite recombination properties. In parallel, minisatellites have been put forward as potential biomarkers for the monitoring of genotoxic agents such as ionizing radiation. We summarize and discuss recent observations on minisatellites. In addition we take advantage of recent whole chromosome sequence data releases to provide a unifying view which may facilitate the annotation of tandem repeat sequences.

Classic Definition and Early Applications of Minisatellites

Minisatellites are usually defined as the repetition in tandem of a short (6- to 100-bp) motif spanning 0.5 kb to several kilobases. Although the first examples described 20 years ago were of human origin, (Wyman and White 1980), similar DNA structures have been found in many organisms including bacteria. Comparisons of the repeat units in classical minisatellites led early on to the notion of consensus or core sequences, which exhibit some similarities with the χ sequence of λ phage (GCTGTGG). In general, the majority of classical minisatellites are GC rich, with a strong strand asymmetry.

Because of their length polymorphism, which results from variations in the number of repeats, and the ability of some of these arrays to cross-hybridize with tens of other similar loci throughout the genome, minisatellites have opened the way to DNA fingerprinting for individual identification (Jeffreys et al. 1985). Minisatellites also provided the first highly polymorphic, multiallelic markers for linkage studies (Nakamura et al. 1987). The usefulness of polymorphic minisatellites (also called VNTRs for variable number of tandem repeats) in the early stages of human genome mapping is reflected in the Centre d'Etude du Polymorphisme Humain/National Institutes of Health consortium linkage maps (National Institutes of Health/Centre d'Etude du Polymorphisme Humain collaborative mapping group 1992).

In parallel, tandem repeats belonging to the minisatellite class were found to be associated with many interesting features of human genome biology and

evolution, usually revealed by pathologies of genetic origin. In brief, minisatellites are thought to contribute to genome function in one of three ways: (1) Some are part of an open reading frame, which may or may not display polymorphism in the human population (for review, see Bois and Jeffreys 1999). (2) Some bind proteins with a variety of functional consequences, strongly suspected or still very hypothetical. Minisatellites located in the 5' region of genes participate in the regulation of transcription (Kennedy et al. 1995). Others located within introns interfere with splicing (Turri et al. 1995). Minisatellites at imprinted loci are thought to play a role in the imprint control (Chaillet et al. 1995; Neumann et al. 1995). More speculatively, minisatellites have been proposed as intermediates in chromosome pairing initiation in some eukaryote genomes (Ashley 1994; Sybenga 1999), which might be related to their proposed recombinogenic properties (Boan et al. 1998; Wahls and Moore 1998). (3) Finally, minisatellites may constitute chromosome fragile sites (for review, see Sutherland et al. 1998) and have been found in the vicinity of a number of recurrent translocation breakpoints and in the switch recombination site in immunoglobulin heavy chain genes (Brusco et al. 1999). These aspects of minisatellite biology have been reviewed elsewhere and will not be further discussed in this article.

Novel Insights and Applications in Minisatellite Biology

Although the high degree of length polymorphism among minisatellites indicates that they are fast-evolving sequences, most of them are in fact quite stable, and neomutated alleles have been observed only at a few loci. Recent research has focused on identifying these rare hypermutable loci in human and

³**Corresponding author.**
E-MAIL Gilles.Vergnaud@igmors.u-psud.fr; FAX 33-1-69-15-66-78.

other genomes because they seem the most appropriate models to illustrate the mechanisms of minisatellite variability. Newly mutated alleles at human hypermutable minisatellites have been characterized in detail, leading to the current model of minisatellite mutation initiation by double-strand breaks (DSBs), and a number of attempts have been made to transfer human minisatellite instability into a more tractable system. We will present and discuss the work done on the subclass of minisatellites that are hypermutable in meiosis.

We will also show how investigations on the sensitivity of minisatellites to some genotoxic agents might provide new insight on minisatellite mutation processes. This work may lead to new applications for minisatellite sequences, such as the development of genotoxicity assays to specifically detect agents interfering with DNA recombination or replication.

Finally, the release of whole genome sequence data allows new approaches to minisatellite characterization. In spite of the fact that our understanding of minisatellite biology has improved very significantly in the last 10 years, minisatellites are usually not discussed or even annotated in releases of new sequence data. This is likely due to the lack of a clear and satisfying definition of these structures. We will briefly review the history of minisatellite characterization and chromosomal localization and compare the picture that these earlier investigations produced to the global view provided now by the sequencing of the human genome.

Insights from the Study of Mutant Alleles at Human Hypermutable Minisatellites

For practical reasons linked to the size of available pedigrees, a minisatellite will usually be classified as hypermutable if its average mutation rate in the germline is higher than 0.5% (the ratio of mutation events in the male and female germline is variable; it can be highly skewed toward paternal events as in CEB1, or equal as in MS1, see Table 1). As a rough estimate, approximately 300 human minisatellites have been typed across families (Armarger et al. 1998; Armour et al. 1990; Nakamura et al. 1987) and less than ten of these qualify as hypermutable (Table 1). The structural features of hypermutable minisatellites described in Table 1 are not specific for this subclass of tandem repeats, and the proportion of telomeric versus interstitial loci (MS32 and MS1 being interstitial) in this collection fits with the proportion of telomeric and interstitial loci among human minisatellites in general (see below).

All hypermutable minisatellites characterized so far possess internal variants, which have provided one way to undertake mutant allele analysis (Table 1). Jeffreys and colleagues developed a polymerase chain re-

action-based assay (Jeffreys et al. 1991) which has proved very efficient at typing the position of variants along alleles. These internal maps can be used to identify the origin of additional repeats in mutant alleles as compared to their progenitors. An important part of our current knowledge of hypermutable minisatellite biology comes from this technology. Two reports in which neomutated alleles at the CEB1 and MS32 hypermutable minisatellites were typed (Buard and Vergnaud 1994; Jeffreys et al. 1994) pointed to DSBs as initiating events of the meiotic mutations. Both interallelic (gene conversion-like) and intra-allelic exchanges were observed, with a different proportion of the two classes of events at the two loci. The detailed typing achieved by the CEB1 study provided data showing that some of the interallelic insertions are flanked by duplicated motifs from the recipient allele. Figure 1 illustrates a model which fits with our current knowledge of meiotic DSBs within hotspots (i.e., in yeast they occur outside the tandem array (Debrauwère et al. 1999) and are almost blunt) while being compatible with observations on CEB1 in the human context.

Subsequent studies have investigated the role of the flanking sequence in the mutation process. This interest in flanking sequences was prompted by the observation that a point mutation very close to one end of the MS32 array was associated with a strongly reduced mutation rate of the corresponding allele (Monckton et al. 1994). In addition, meiotic mutation events in MS32 strongly clustered toward one end of the array (Jeffreys et al. 1994). In contrast, somatic mutations at MS32 (Jeffreys and Neumann 1997) do not show clustering toward one end; they occur at a much lower frequency and are simple intra-allelic events, predominantly deletions.

Experimental Models of Minisatellite Mutation

The development of experimental models to study minisatellite mutation processes has been necessary in order to analyze more precisely the timing of the mutation processes, the underlying genetics, and test the predictions made by the current models.

Attempts to develop animal models based on the identification of naturally occurring hypermutable minisatellites failed. Two hypermutable tandem repeats characterized in mice are the amplification of short (respectively, 4- and 5-bp) units, which do not fully qualify as minisatellites and are not amenable to variant typing (for review, see Bois and Jeffreys, 1999). For this reason, Jeffreys and colleagues developed a transgenic mouse model. The inserts injected were either an MS32 tandem array with only a few hundred base pairs of flanking sequence or a complete cosmid insert from the MS32 or CEB1 locus. Interestingly, although the mitotic instability was transferred, no mei-

Table 1. Known Human Hypermutable Minisatellites

Minisatellite name	Locus or D-number	Location	Accession number	Consensus length	Nb of variant positions	Percent match (%M)	GC% (bias)	Purine % (bias)	Instability	Proportion of paternal events
CEB1	D2S90	2q37.3 (T)	AF048727	39	8	90	72 (0.53)	68 (0.38)	6.7%	97%
CEB15	D1S172	1p36.33 (T)	AL096805	18	8	91	68 (0.33)	77 (0.56)	1.5%	100%
CEB25	D10S180	10q26.3 (T)	AL096806	52	>20	82	53 (0.73)	67 (0.36)	2.5%	65%
CEB36	D10S473	10q26.3 (T)	AL096810	42	18	93	64 (0.65)	66 (0.32)	1.8%	50%
CEB42	D8S358	8q24.3 (T)	AL096807	41	4	95	63 (0.11)	53 (0.08)	0.5%	100%
CEB72	D17S888	17q25 (T)	AL096808	21	0	100	71 (0.60)	80 (0.62)	1.8%	65%
MS1	D1S7	1p33-35 (I)	X14856	9	3	85	58 (0.93)	79 (0.60)	5%	50%
MS32	D1S8	1q42 (I)	AF048729	29	3	92	61 (0.37)	67 (0.35)	0.4%	50%
B6.7		20q13.3 (T)	AF081787	34	6	91	60 (0.5)	68 (0.37)	5.5%	64%
CSTB	EPM1	21q22.3 (T)	U46692	12	1	96	100 (0.5)	100 (0.5)	47%	76%

Minisatellites showing instability higher than 1% in the male or female germline are listed (with the exception of MS32, with a lower mutation rate, but which is a reference minisatellite in many investigations). The instability values indicated are most often average values, as measured usually in the large Centre d'Etudes du Polymorphisme Humain families (<http://www.cephb.fr>) with the exception of the cystatin B (CSTB) gene minisatellite. In this last case, the values given were measured in pathogenic, expanded alleles (Larson et al. 1999). Similarly, the mutation rate at CEB1 alleles has been shown to vary between <0.02% (at smaller alleles) and >20% (Buard et al. 1998). Percent match (%M) is the average similarity of motifs with the consensus motif. GC bias is the absolute value of (G% - C%)/(G% + C%). Purine bias is the absolute value of (pur% - pyr%).

otic instability was observed (Bois et al. 1997; Buard et al. 2000). However, in these investigations the integration site was random, and no attempts were made to target potentially more active loci of the mouse genome.

Alternative approaches have used yeast. The work in yeast was pioneered by Rannug, Cederberg, and colleagues, who showed meiotic induction of human minisatellite MS32 instability. The minisatellite was inserted in the vicinity of the LEU2 yeast hotspot for recombination initiation, where DSBs frequently form (Appelgren et al. 1997). Tetrad analysis demonstrated that interallelic mutants, which might look like bona fide crossover events (exchange of flanking markers, no complex secondary rearrangements), are in fact conversion events (Appelgren et al. 1999), which is of

some importance when interpreting similar human data (Jeffreys et al. 1998). In a more recent investigation, the similar meiotic instability of a CEB1 allele introduced in yeast was shown to be dependent on the integration site (Debrauwère et al. 1999). Integration in a cold spot for recombination initiation resulted in a very low meiotic instability compared to integration adjacent to the ARG4 recombination hotspot. At this site, the tandem array did not modify the DSB hotspot: DSBs remained detectable on both sides of the array at a frequency comparable to the wild-type situation. Suppression of the DSBs, either by a failure in activating the site, as obtained in a rad50 deficient strain, or by the absence of the topoisomerase (Spo11) responsible for the DSBs (Bergerat et al. 1997), reduced the meiotic instability of the minisatellite to the mitotic level. Finally, taking advantage of mismatch repair-deficient strains, the predicted heteroduplex intermediates (Fig. 1) have been observed in some (but not all) mutant alleles.

These observations, combined with the attempts to develop a mouse model and the data in humans, very strongly suggest that the production of an experimental model in which the minisatellite shows meiotic

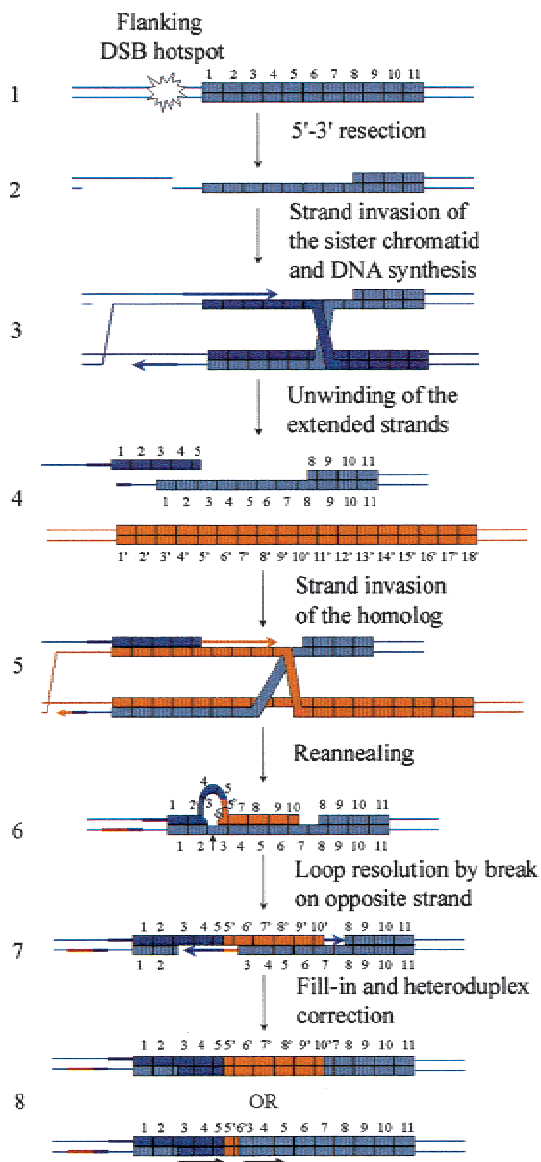


Figure 1 Revised model for meiotic mutation events demonstrating the formation of interallelic events with duplications flanking the converted motifs. In order to explain the observed duplication flanking meiotic interallelic events, the model initially proposed by Buard and Vergnaud (1994) and subsequently adopted by others (Bois and Jeffreys 1999) invoked DSBs initiated within the array by staggered single-strand breaks separated by 80 nucleotides or more. This would require a strong associated helicase activity and does not fit with the view now provided by the yeast work (Debrauwère et al. 1999). Alternatively, the presented model adapted from Debrauwère et al. (1999) shows how an almost blunt DSB, produced in a flanking DSB hotspot outside the minisatellite (step 1), can produce interallelic exchanges with a duplication flanking the converted motifs, as well as most, if not all, minisatellite rearrangements observed in man or yeast. After 5'-3' resection (step 2), the repair is initiated by invading the sister chromatid and priming DNA synthesis on one or both (as suggested here, step 3) strands. After DNA synthesis, the newly synthesized strands independently unwind (step 4) and are free to engage in other DNA-DNA interactions (here, both strands are shown invading the homolog, step 5). Eventually, the newly synthesized strands reanneal together with properly aligned flanking sequences. A loop may form on one (as shown here, step 6) or both (Debrauwère et al. 1999) strands. This loop can be converted into the corrected allele (Debrauwère 2000) via a single strand cut on the opposing DNA strand (step 7), or removed. Depending on which strand is used to correct the heteroduplex, a direct duplication of repeats flanking the converted patch is produced (step 8, bottom). All models proposed so far predict the existence of patches of heteroduplex intermediates produced by the reannealing of similar, but different, minisatellite motifs (here, step 6). This last prediction was successfully tested in Debrauwère et al. (1999). An interesting aspect of this model is that the lower strand may extend in the flanking sequence at steps 2 and 4. This will produce a heteroduplex region in the flanking sequence (steps 6 and 7; left unrepaired here in step 8) which, once repaired, may introduce a conversion patch in the final product flanking sequence.

instability depends on the coincidence of a tandem repeat with a DSB hotspot. Conversely, minisatellites can be made unstable in mitosis in yeast strains deficient for some aspects of DNA replication (Kokoska et al. 1998), so the yeast model has already provided experimental support for the view that at least two mechanisms promote minisatellite instability (Jeffreys and Neumann 1997). Accordingly, it can be anticipated that agents interfering with one of these mechanisms will induce minisatellite instability.

Genotoxicity

A number of studies indicate that hypermutable minisatellites might provide biomarkers for exposure to some genotoxic agents. One such class of genotoxic agents is ionizing radiation. The first hint of the sensitivity of minisatellites to ionizing radiation was obtained in mouse (Dubrova et al. 1993; Dubrova et al. 1998). This was followed by studies in humans exposed to chronic low doses of radiation based on populations living in regions contaminated by the release of radioactive material after the explosion at the Chernobyl power station in 1986. The investigation used Southern blotting to genotype father-mother-child trios at hypermutable minisatellite loci and to count the frequency of mutant alleles in a control and an exposed population (Dubrova et al. 1997). The data obtained indicated that the frequency of mutant alleles in the exposed population was twice the frequency observed in the control population (from the United Kingdom). Importantly, the exposed population was split into two parts according to the degree of soil contamination in regions from which families were collected, suggesting a dose-effect relationship. The results are in contrast with the Hiroshima-Nagasaki survivors investigations (Sato and Kodaira 1996), but the situation in the Hiroshima-Nagasaki study is very different because children were conceived years after parental exposure. At this time, minisatellite mutation rate in the germline should be back to normal, if data obtained with the mouse model can be extrapolated to human (Dubrova et al. 1998).

Several chemicals released in the environment are also suspected of inducing meiotic minisatellite mutations. Germline mutation rate monitored by DNA fingerprinting was twice as high in herring gulls inhabiting a heavily industrialized area as compared to birds living in rural sites (Yauk and Quinn 1996). Similarly, instability of the human minisatellite MS32 introduced in yeast also appears to be modulated by some chemicals (Appelgren et al. 1999).

Taking Advantage of the Global View Provided by Large-Scale Sequencing

During the 1980s and early 1990s, a number of approaches were developed to detect and/or clone mini-

satellite loci. Because DNA fingerprinting, using so-called multilocus minisatellite probes, previously demonstrated the property of some tandem arrays to cross-hybridize with a number of others, the majority of these approaches was based on cross-hybridization detection (Vergnaud 1989). Given the technology which was used, i.e., Southern blotting, a minisatellite would be defined as a tandem repeat with allele length usually in the range that can be assayed by Southern blots, i.e., above approximately 800 bp.

The overall frequency of such minisatellites in five mammalian genomes investigated at a significant scale is similar (Amarger et al. 1998; Bois et al. 1998; Georges et al. 1991). The distribution is however very different, with a high bias toward chromosome ends in human and a much lower bias in mouse and rat. The situation in the pig is intermediate, and a closer look at the synteny relationships suggests that, in a common ancestor, the interstitial minisatellite clusters were telomeric (Amarger et al. 1998). One conclusion of these investigations is that the tandem repeats which can be analyzed on a Southern blot are predominantly associated with chromosome ends, and internal clusters of such tandem repeats are very likely to be the result of secondary rearrangements such as chromosome ends fusion.

However, analyses limited to the usual definition of minisatellites (>800 bp) are not altogether satisfying because this definition represents only a fraction of tandem repeats, many of which are smaller than the 500-bp arbitrary limit, but do not fit in the microsatellite class of tandem repeats. Furthermore, this definition has a limited value when dealing with sequence data for at least two reasons: 1) tandem repeats which clearly qualify as minisatellites often have some alleles in the human population which are much shorter than the 500-bp limit, and 2) during the assembly of raw sequence data, the true allele length of minisatellites is not always correctly inferred, especially when the internal array is very homogeneous. CEB1, the most hypermutable minisatellite characterized so far (Vergnaud et al. 1991), illustrates both of these drawbacks of the current definition: (1) Small alleles with 5 repeat units (total array length: 200 bp) have been described, and their meiotic mutation rate is still high at 0.4% (Buard et al. 1998). (2) The cosmid from which CEB1 was originally isolated has been sequenced (Murray et al. 1999). Although the CEB1 allele present in this cosmid is 3.6 kb long, as estimated by restriction enzyme analysis, the deposited cosmid sequence contains only six CEB1 motifs spanning 240 bp, presumably because of difficulties encountered in sequencing the array.

The release of whole chromosome sequence data for a number of eukaryotes including human, the nematode *Caenorhabditis elegans*, and the plant *Arabidopsis thaliana* now opens the way to more systematic,

sequence-based investigations. For this purpose, we have constructed a prototype tandem repeat database (<http://minisatellites.u-psud.fr>) using the Tandem Repeats Finder software (Benson 1999) to identify the repeats. The database contains more than 14,000 tandem repeats for the acrocentric chr22 (34.6 Mb) (Dunham et al. 1999) and can be queried according to a number of features (see legend, Fig. 2). One-third of human chr22 tandem repeats (Fig. 2A) satisfies an enlarged definition of minisatellites, as used in this review (at least three units, repeat unit longer than 6 bp). Among them, minisatellites with repeat units longer than 16 bp and total length greater than 100 bp display a distribution strongly biased toward the chr22 long-arm telomere (Fig. 2B).

Figure 2C shows the result of a query mimicking characteristics of classical minisatellites, i.e., query "B" plus a high GC content, strong strand bias, and strong

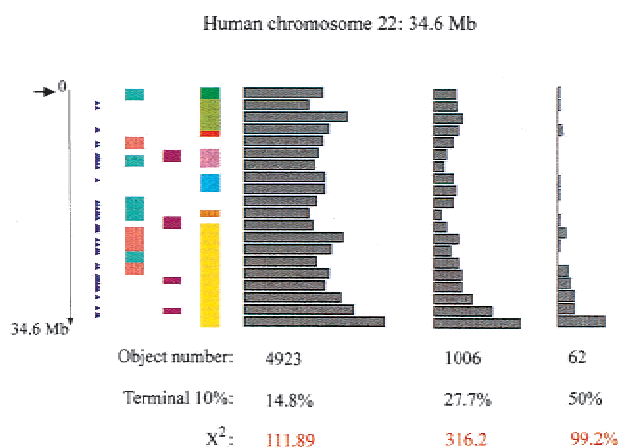


Figure 2 Distribution of tandem repeats corresponding to different queries along human chromosome 22. Tandem repeats have been identified within the human chromosome 22 sequence using the Tandem Repeats Finder (TRF) software with the following options: alignment parameters = (2,3,5), minimum alignment score to report repeat = 50, maximum period size = 500. Redundancy was then eliminated, and Alu and satellite sequences (152 were identified) were filtered. The arrow (top left) shows the centromere position. The position of the 51 chr22 Genethon microsatellites present in the database is shown with arrow heads. GC-rich (pink) or -poor (green) areas, regions of increased recombination, and known mouse synteny correspondence are as indicated in Dunham et al. (1999). Distributions obtained with different queries: (A, Left) $U \geq 6, N \geq 3$ (B, Middle) $U > 16, N \geq 3, L > 100$ (C, Right) $U > 16, N \geq 3, \%GC \geq 65\%, BGC \geq 0.3, \%M \geq 85\%$ (U = unit length, N = copy number, L = total length, $\%GC$ = GC percent, BGC = G/C bias = $|\%G - \%C| / (\%G + \%C)$, $\%M$ (percent matches) is the average similarity of each motif with the consensus motif). Percentages reported correspond to the proportion of objects in the last 10% of total length. χ^2 values were calculated by comparing the last 10% of the chromosome with the mean number of objects along the whole chromosome. χ^2 threshold of significance (homogeneity hypothesis is rejected if χ^2 is greater than threshold) is 3.841 with $P = 5\%$, and 10.827 with $P = 0.1\%$ (1 degree of freedom).

internal homogeneity (see legend, Fig. 2, for details). Half of the 62 minisatellites fitting this query are located within the terminal 10% of chr22. Such simple queries demonstrate that a fraction of minisatellites, comprising hundreds of loci on chromosome 22 alone, do behave as initially suggested by the subset of classical minisatellites, i.e., are present at a much higher frequency in the terminal R band of human chromosomes (Amarger et al. 1998).

Chromosome ends appear to be relatively poor in recombination nodules during human male meiosis, which is surprising given the very high male recombination rates observed toward chromosome ends. Subtelomeric minisatellites are one class of sequences that have been put forward as candidates to help explain this paradox (Ashley 1994; Sybenga 1999). Specific mechanisms would be activated in male meiosis, and minisatellites would be involved in chromosome pairing, either directly or via interactions with pairing proteins. This predicts that minisatellites should not display subtelomeric clustering in plants, where no such discrepancy between recombination nodules and rates is observed. Figure 3 presents comparisons of the three species using *C. elegans* chromosome 1 (12.75 Mb) and *A. thaliana* chromosome 4 (17.8 Mb) (The *C. elegans* Sequencing Consortium 1998; Mayer et al. 1999). The total number of tandem repeats found with Tandem Repeats Finder in the three species is not proportional to chromosome length (Fig. 3). It is significantly higher in the nematode (637 Mb) when compared to man and *A. thaliana* (415 Mb and 445 Mb, respectively).

The result of a representative query is shown in Figure 3, bottom row. The number of positive minisatellites is similar in the three species, taking into account chromosome size difference. A strong telomeric bias is observed for *C. elegans* chr1, (right panel) reminiscent of the situation in human chr22. In contrast, the distribution of minisatellites in *A. thaliana* (middle) is strikingly different from that of the two other genomes: tandem repeats are mainly located around the centromere. Figure 4A plots, for each species, the ratio of telomeric versus nontelomeric tandem repeats according to repeat unit length. *C. elegans* chr1 demonstrates telomeric bias for both short units (in particular, 6- and 12-bp units, due to the presence of many (TTAGGC) n telomere-like tandem arrays (The *C. elegans* Sequencing Consortium 1998) and longer units (above approximately 18 bp). Human chr22 demonstrates telomeric bias for repeat units above 17 bp. It may be worth noting that in yeast, 16 bp is the threshold above which mismatch repair mechanisms are unable to correct DNA loops (Sia et al. 1997). Figure 4B plots the same measure of telomeric bias according to the overall array length. In contrast with *C. elegans*, the telomeric bias for human chr22 appears only for arrays longer than 120–140 bp. This threshold is reminiscent

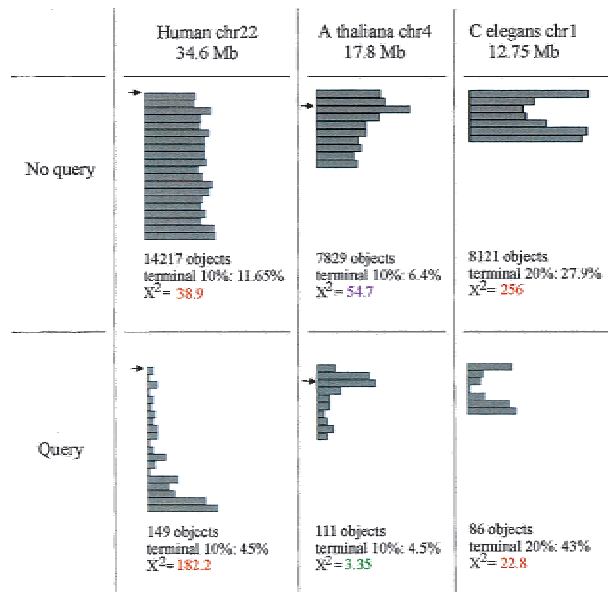


Figure 3 Comparison of tandem repeats distribution in three species. See legend, Fig. 2, for information on database construction and other details. Arrows show centromere position (unknown for nematode). The Tandem Repeats Finder software identifies tandem repeats at a frequency of 415 per Mb for human chromosome 22, 445 per Mb for *Arabidopsis thaliana* chromosome 4, and 637 per Mb for *Caenorhabditis elegans* chromosome 1 ("No query" panel). Bottom panel: the query applied was $U \geq 10$, $N \geq 3$, $L > 100$, $0.3 \geq \text{BGC} \geq 0.55$, $\%M \geq 70$. Chromosomes were fragmented in areas of comparable length: 1.73 Mb (20 areas), 1.78 Mb (10 areas), and 1.82 Mb (7 areas), for human chromosome 22, plant chromosome 4 and nematode chromosome 1, respectively. For human chromosome 22 and nematode chromosome 1, significant telomeric biases are observed. In contrast, the plant chromosome shows a bias toward the centromeric region.

of triplet repeat instability observed above 40–50 repeats. No telomeric bias is observed in *A. thaliana* chr4.

Concluding Remarks and Perspectives

Previously, the number of classical minisatellites has been estimated to be a few thousand in the human genome, which translates to a few tens on chromosome 22. Such rare objects would not likely play a significant role in genome metabolism. The view now provided by the availability of whole human chromosome sequence reveals a much larger number of small minisatellites with repeat units similar to the classical structures and a similarly biased distribution toward chromosome ends, which is not observed in *A. thaliana*. These observations give much more credibility to these structures (Boan et al. 1998; Wahls and Moore 1998). Obviously, comparisons with additional, larger human chromosomes will be of some interest.

It is tempting to speculate that the meiotic hypermutability of some minisatellite structures is the by-product of the coincidence of an ordinary minisatellite

with a DSB hotspot (Debrauwère et al. 1999). The disappearance of a hotspot, as proposed by Boulton et al. (1997) will then remove the hypermutability of the neighboring tandem repeat. In this model, the study of hypermutable minisatellites is demonstrating more about human DSB hotspots, the majority of which would exist independently of neighboring tandem repeats in human (Badge et al. 2000) as in yeast, than about minisatellites in general. The model presented in Figure 1 shows how a double strand break occurring outside of the array (as suggested by Debrauwère et al. 1999) can indeed produce the complex interallelic events observed in man, including duplications flanking the converted patch. The model also accommodates conversion patches in the flanking sequence, which may include mosaics of intra- and interallelic origin. In contrast, the making of minisatellites in general would result from replication mechanisms, favored by deficiencies in enzymes involved in replication such as *Saccharomyces cerevisiae* Rad27 as proposed in Tishkoff et al. (1997). In the process, sequence features of the motif, likely to produce secondary structures or slow down the polymerase on the lagging strand during replication (G-rich DNA strands, palindromic motifs in AT rich minisatellites, GC richness), may be important.

In this regard, no information regarding minisatellite instability or even polymorphism is obtained using the tandem repeat database presented here. This will be an important further step of the database development, which might benefit from the current knowledge of variant motif interspersions patterns along hypermutable minisatellite alleles. In addition, tandem repeat polymorphism predictions will be facilitated by the expected availability, in the near future, of sequence data from more than one allele.

Genotoxicity is a promising domain for minisatellite-related investigation. It may combine short-term applications toward the development of genotoxicity assays specifically identifying recombinogenic activities with more basic investigations into the purpose of minisatellites and what triggers them. One question raised by these investigations is whether the tandem array itself is the target of the genotoxic agent, whether it is the flanking DSB hotspot which is further activated by the agent, or whether it is the replication machinery which is affected. In the second hypothesis, hypermutable minisatellites would act as markers for the activity of their flanking recombination hotspot, whereas in the first (and perhaps also third) hypothesis, any minisatellite could act as a biomarker for the genotoxic agent. Recently developed yeast models may help address such issues.

ACKNOWLEDGMENTS

We thank Christine Pourcel for comments and critical reading

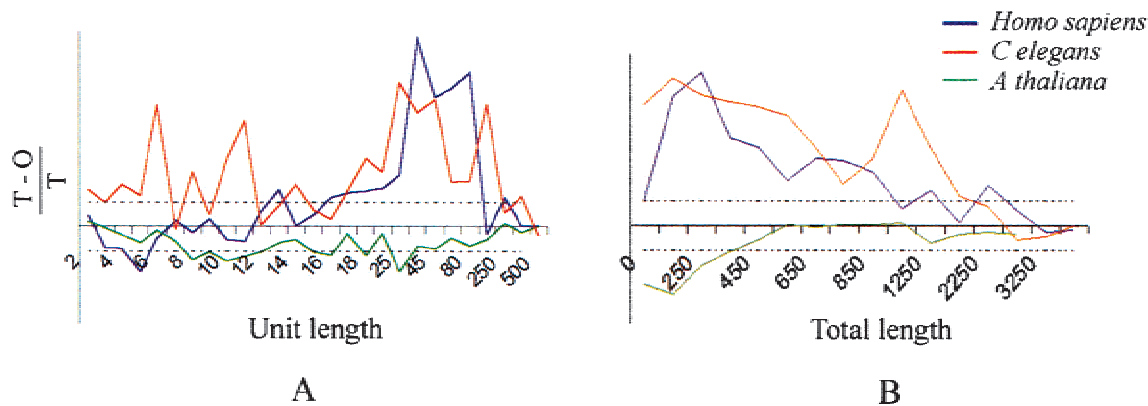


Figure 4 Comparison between terminal and other regions according to unit length (A) or total length (B). On Y-axis: {number of objects in the terminal 10% of the sequence [T] – number of object in other regions [mean for 10%] [O]}/O (corresponds to a Z-score). If >0, terminal 10% are richer than the rest of the genome; else they are poorer, with a significance threshold of 1.96 (dotted lines). On X-axis: unit length (A) or total array length (B).

of this work. Current minisatellite work in the laboratory is supported by a grant from Délégation Générale de l'Armement (DGA/DSP/STTC).

REFERENCES

- Amarger, V., Gauguier, D., Yerle, M., Apiou, F., Pinton, P., Giraudeau, F., Monfouilloux, S., Lathrop, M., Dutrillaux, B., Buard, J. et al. 1998. Analysis of the human, pig, and rat genomes supports a universal telomeric origin of minisatellite sequences. *Genomics* **52**: 62–71.
- Appelgren, H., Cederberg, H., and Rannug, U. 1997. Mutations at the human minisatellite MS32 integrated in yeast occur with high frequency in meiosis and involve complex recombination events. *Mol. Gen. Genet.* **256**: 7–17.
- Appelgren, H., Cederberg, H., and Rannug, U. 1999. Meiotic interallelic conversion at the human minisatellite MS32 in yeast triggers recombination in several chromatids. *Gene* **239**: 29–38.
- Appelgren, H., Hedenskog, M., Sandstrom, C., Cederberg, H., and Rannug, U. 1999. Polychlorinated biphenyls induce meiotic length mutations at the human minisatellite MS32 in yeast. *Environ. Mol. Mutagen* **34**: 285–290.
- Armour, J.A.L., Povey, S., Jeremiah, S., and Jeffreys, A.J. 1990. Systematic cloning of human minisatellites from ordered array charomid libraries. *Genomics* **8**: 501–512.
- Ashley, T. 1994. Mammalian meiotic recombination: a reexamination. *Hum. Genet.* **94**: 587–593.
- Badge, R.M., Yardley, J., Jeffreys, A.J., and Armour, J.A. 2000. Crossover breakpoint mapping identifies a subtelomeric hotspot for male meiotic recombination. *Hum. Mol. Genet.* **9**: 1239–1244.
- Benson, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**: 573–580.
- Bergerat, A., de Massy, B., Gabelle, D., Varoutas, P.-C., Nicolas, A., and P. Forterre. 1997. An atypical topoisomerase II from archaea with implication for meiotic recombination. *Nature* **386**: 414–417.
- Boan, F., Rodriguez, J.M., and Gomez-Marquez, J. 1998. A non-hypervariable human minisatellite strongly stimulates in vitro intramolecular homologous recombination. *J. Mol. Biol.* **278**: 499–505.
- Bois, P., Collick, A., Brown, J., and Jeffreys, A.J. 1997. Human minisatellite MS32 (D1S8) displays somatic but not germline instability in transgenic mice. *Hum. Mol. Genet.* **6**: 1565–1571.
- Bois, P. and Jeffreys, A.J. 1999. Minisatellite instability and germline mutation. *Cell Mol. Life Sci.* **55**: 1636–1648.
- Bois, P., Stead, J.D., Bakshi, S., Williamson, J., Neumann, R., Moghadaszadeh, B., and Jeffreys, A.J. 1998. Isolation and characterization of mouse minisatellites. *Genomics* **50**: 317–330.
- Boulton, A., Myers, R.S., and Redfield, R.J. 1997. The hotspot conversion paradox and the evolution of meiotic recombination. *Proc. Natl. Acad. Sci.* **94**: 8058–8063.
- Brusco, A., Saviozzi, S., Cinque, F., Bottaro, A., and DeMarchi, M. 1999. A recurrent breakpoint in the most common deletion of the Ig heavy chain locus (del A1-GP-G2-G4-E). *J. Immunol.* **163**: 4392–4398.
- Buard, J., Bourdet, A., Yardley, J., Dubrova, Y., and Jeffreys, A.J. 1998. Influences of array size and homogeneity on minisatellite mutation. *EMBO J.* **17**: 3495–3502.
- Buard, J., Collick, A., Brown, J., and Jeffreys, A.J. 2000. Somatic versus germline mutation processes at minisatellite CEB1 (D2S90) in humans and transgenic mice. *Genomics* **65**: 95–103.
- Buard, J. and Vergnaud, G. 1994. Complex recombination events at the hypermutable minisatellite CEB1 (D2S90). *EMBO J.* **13**: 3203–3210.
- Chaillet, J.R., Bader, D.S., and Leder, P. 1995. Regulation of genomic imprinting by gametic and embryonic processes. *Genes Dev.* **9**: 1177–1187.
- Debrauwère, H. 2000. Analyse des mécanismes d'instabilité des séquences répétées humaines de type minisatellite dans la levure *S. cerevisiae*. *Biologie-Sciences de la vie*. PhD thesis, University Paris VI: 56–61.
- Debrauwère, H., Buard, J., Tessier, J., Aubert, D., Vergnaud, G., and Nicolas, A. 1999. Meiotic instability of human minisatellite CEB1 in yeast requires DNA double-strand breaks. *Nat. Genet.* **23**: 367–371.
- Dubrova, Y.E., Jeffreys, A.J., and Malashenko, A.M. 1993. Mouse minisatellite mutations induced by ionizing radiation. *Nat. Genet.* **5**: 92–94.
- Dubrova, Y.E., Nesterov, V.N., Krouchinsky, N.G., Ostapenko, V.A., Vergnaud, G., Giraudeau, F., Buard, J., and Jeffreys, A.J. 1997. Further evidence for elevated human minisatellite mutation rate in Belarus eight years after the Chernobyl accident. *Mut. Res.* **381**: 267–278.
- Dubrova, Y.E., Plumb, M., Brown, J., Fennelly, J., Bois, P., Goodhead, D., and Jeffreys, A.J. 1998. Stage specificity, dose response, and doubling dose for mouse minisatellite germ-line mutation induced by acute radiation. *Proc. Natl. Acad. Sci.* **95**: 6251–6255.
- Dunham, I., Shimizu, N., Roe, B.A., Chisoo, S., Hunt, A.R., Collins, J.E., Bruskiewich, R., Beare, D.M., Clamp, M., Slink, L.J. et al. 1999. The DNA sequence of human chromosome 22. *Nature* **402**: 489–495.
- Georges, M., Gunawardana, A., Threadgill, D.W., Lathrop, M., Olsaker, I., Mishra, A., Sargeant, L.L., Schoeberlein, A., Steele, M.R., Terry, C. et al. 1991. Characterization of a set of variable number of tandem repeat markers conserved in bovidae. *Genomics* **11**: 24–32.

- 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. *Nature* **354**: 204–209.
- Jeffreys, A.J., Neil, D., and Neumann, R. 1998. Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J.* **17**: 4147–4157.
- Jeffreys, A.J. and Neumann, R. 1997. Somatic mutation processes at a human minisatellite. *Hum. Mol. Genet.* **6**: 129–136.
- 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. *Nat. Genet.* **6**: 136–145.
- Jeffreys, A.J., Wilson, V., and Thein, S.L. 1985. Individual-specific 'fingerprints' of human DNA. *Nature* **316**: 76–79.
- Kennedy, G.C., German, M.S., and Rutter, W.J. 1995. The minisatellite in the diabetes susceptibility locus IDDM2 regulates insulin transcription. *Nat. Genet.* **9**: 293–298.
- Kokoska, R.J., Stefanovic, L., Tran, H.T., Resnick, M.A., Gordenin, D.A., and Petes, T.D. 1998. Destabilization of yeast micro- and minisatellite DNA sequences by mutations affecting a nuclease involved in Okazaki fragment processing (*rad27*) and DNA polymerase delta (*pol3-t*). *Mol. Cell Biol.* **18**: 2779–2788.
- Larson, G.P., Ding, S., Lafreniere, R.G., Rouleau, G.A., and Krontiris, T.G. 1999. Instability of the EPM1 minisatellite. *Hum. Mol. Genet.* **8**: 1985–1988.
- Mayer, K., Schuller, C., Wambutt, R., Murphy, G., Volckaert, G., Pohl, T., Dusterhoft, A., Stiekema, W., Entian, K.D., Terryn, N. et al. 1999. Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* **402**: 769–777.
- Monckton, D.G., Neumann, R., Guram, T., Fretwell, N., Tamaki, K., MacLeod, A., and Jeffreys, A.J. 1994. Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nat. Genet.* **8**: 162–170.
- Murray, J., Buard, J., Neil, D.L., Yeramian, E., Tamaki, K., Hollies, C.R., and Jeffreys, A.J. 1999. Comparative sequence analysis of human minisatellites showing meiotic repeat instability. *Genome Res.* **9**: 130–136.
- Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. et al. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**: 1616–1622.
- Neumann, B., Kubicka, P., and Barlow, D.P. 1995. Characteristics of imprinted genes. *Nat. Genet.* **9**: 12–13.
- NIH/CEPH Collaborative Mapping Group. 1992. A comprehensive genetic linkage map of the human genome. *Science* **258**: 67–83.
- Satoh, C. and Kodaira, M. 1996. Effects of radiation on children. *Nature* **383**: 226.
- Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P., and Petes, T.D. 1997. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell Biol.* **17**: 2851–2858.
- Sutherland, G.R., Baker, E., and Richards, R.I. 1998. Fragile sites still breaking. *Trends Genet.* **14**: 501–506.
- Sybenga, J. 1999. What makes homologous chromosomes find each other in meiosis? A review and an hypothesis. *Chromosoma* **108**: 209–219.
- The *C. elegans* Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**: 2012–2018.
- Tishkoff, D.X., Filosi, N., Gaida, G.M., and Kolodner, R.D. 1997. A novel mutation avoidance mechanism dependent on *S. cerevisiae* *RAD27* is distinct from DNA mismatch repair. *Cell* **88**: 253–263.
- Turri, M.G., Cuin, K.A., and Porter, A.C. 1995. Characterisation of a novel minisatellite that provides multiple splice donor sites in an interferon-induced transcript. *Nucleic Acids Res.* **23**: 1854–1861.
- Vergnaud, G. 1989. Polymers of random short oligonucleotides detect polymorphic loci in the human genome. *Nucleic Acids Res.* **17**: 7623–7630.
- 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.
- Wahls, W.P. and Moore, P.D. 1998. Recombination hotspot activity of hypervariable minisatellite DNA requires minisatellite DNA binding proteins. *Somat. Cell Mol. Genet.* **24**: 41–51.
- Wyman, A. R. and White, R. 1980. A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci.* **77**: 6754–6758.
- Yauk, C.L. and Quinn, J.S. 1996. Multilocus DNA fingerprinting reveals high rate of heritable genetic mutation in herring gulls nesting in an industrialized urban site. *Proc. Natl. Acad. Sci.* **93**: 12137–12141.