

ORIGINAL ARTICLE

Tsunenori Mizukoshi · Keiji Tamaki · Jun-ichi Azumi
Hiroshi Matsumoto · Kohzoh Imai · Alec J. Jeffreys

Allelic structures at hypervariable minisatellite B6.7 in Japanese show population specificity

Received: January 29, 2002 / Accepted: February 12, 2002

Abstract Human minisatellite B6.7 shows extensive allele length and structural variability in north Europeans. We analysed this locus in the Japanese population. Allele size distributions showed that Japanese retain extensive allele length variability but have significantly smaller alleles compared with north Europeans. In contrast, there is very little variation in flanking DNA, with only one single-nucleotide polymorphism (SNP) near the minisatellite. Ninety-two Japanese alleles were further characterised by minisatellite variant repeat mapping by polymerase chain reaction (MVR-PCR). These alleles showed a wide variety of internal MVR structures, despite their relative shortness, with most alleles observed only once in the sample. The true heterozygosity is estimated at 99.95%, with well in excess of 2000 different alleles existing in the Japanese population. Dot matrix analysis showed that groups of related alleles sharing structural motifs could be identified within Japanese and in north Europeans, and that these groups are population specific with no examples of significant similarity between any Japanese and north European alleles. Minisatellite B6.7 therefore shows huge allele variability and fast repeat turnover in Japanese as well as north European populations, and provides novel lineage markers for exploring very recent events in human population history.

Key words Minisatellite B6.7 · MVR-PCR · VNTR · Mutation · Human diversity

Introduction

Human minisatellites or variable number tandem repeat loci are a class of tandem repetitive DNA comprising about 3% of the human genome (International Human Genome Sequencing Consortium 2001). Human GC-rich minisatellites are preferentially found clustered in the recombination-proficient subtelomeric regions in chromosomes (Royle et al. 1988). The repeat unit length ranges from 6 to more than 100bp, with arrays usually kilobases in length. Some minisatellite loci show very high levels of allele length variability. Most loci consist of heterogeneous arrays of two or more subtly different repeat unit types (minisatellite variant repeats). These variant repeats can be mapped by the polymerase chain reaction (MVR-PCR) to provide a powerful approach for allele classification based on the interspersal patterns of variant repeats within the repeat array (Jeffreys et al. 1991). MVR-PCR followed by agarose gel electrophoresis and Southern blot hybridisation allows such interspersal patterns to be displayed as ladders of PCR products. This approach has revealed enormous levels of allelic variation at several hypervariable minisatellites (Jeffreys et al. 1991; Neil and Jeffreys 1993; Armour et al. 1993; Buard and Vergnaud 1994; Andreassen and Olaisen 1998; Tamaki et al. 1999).

Human minisatellite B6.7 (DDBJ/EMBL/GenBank accession no. AF081787) was discovered as a hypervariable locus in the terminal region of chromosome 20q13.33 (Kimpton et al. 1992). Variation and mutation at this locus was further characterised in north Europeans (Tamaki et al. 1999). Alleles contain from 6 to over 500 repeats, and this variability is maintained by a very high germline mutation rate of 7.0% per gamete in men and 3.9% in women, indicating that B6.7 is one of the most unstable minisatellites isolated to date. Small-pool PCR analysis of sperm and blood DNA showed that repeat instability is restricted to the germline and is most likely meiotic in origin, and that instability increases with tandem array size. Structural analysis of sperm mutants by MVR-PCR revealed a wide variety of changes in allele structure, including simple intra-

T. Mizukoshi (✉) · K. Tamaki · J. Azumi · H. Matsumoto
Department of Legal Medicine, Sapporo Medical University School
of Medicine, South 1, West 17, Chuo-ku, Sapporo 060-8556, Japan
Tel. +81-11-611-2111; Fax +81-11-611-3935
e-mail: tm13@pop01.odn.ne.jp

T. Mizukoshi · K. Imai
First Department of Internal Medicine, Sapporo Medical University
School of Medicine, Sapporo, Japan

A.J. Jeffreys
Department of Genetics, University of Leicester, Leicester, UK

allelic duplications and deletions and more complicated inter-allelic and intra-allelic transfers of repeat blocks, as seen at other human minisatellites (Jeffreys et al. 1994; Buard et al. 1998). There was no clear evidence of mutation polarity at B6.7, as seen at some other minisatellites (Armour et al. 1993; Jeffreys et al. 1994). Some mutants showed extraordinarily complex rearrangements, with evidence of repeated inter-allelic transfers plus the generation of novel repeats by rearrangement at the subrepeat level, suggesting that recombinational instability at B6.7 is a complex multistep process.

Japanese alleles of hypervariable minisatellites such as MS32, MS31A, and MS205 have been characterised previously by MVR-PCR (Jeffreys et al. 1991; Tamaki et al. 1992b; Neil and Jeffreys 1993; Huang et al. 1996; Armour et al. 1996). Japanese alleles showed high levels of variety as well as north European alleles. However, levels of the allele similarity to north Europeans showed difference amongst the loci, possibly depending on the speed of repeat turnover. In the present study, we characterise Japanese alleles at minisatellite B6.7 to explore allele diversity and the population specificity or otherwise of structurally related alleles at this remarkably variable locus.

Subjects and methods

DNA samples

DNA was prepared by standard phenol/chloroform extraction from peripheral leukocytes of 120 healthy unrelated Japanese volunteers with their informed consent. The concentration of genomic DNA sample was quantified fluorometrically using a TKO 100 fluorometer (Hoefer Scientific Instruments, San Francisco, CA, USA). North European DNA samples were as described previously (Tamaki et al. 1999). The present study, including the use of the Japanese samples, was approved by the Ethics Committee of Sapporo Medical University.

Allele length distribution at minisatellite B6.7

Samples of 50–150 ng of blood DNA were amplified in 7- or 10- μ l reactions using the PCR buffer described previously (Jeffreys et al. 1990) plus 0.4 μ M primers 67B and 67C (Tamaki et al. 1999) and 0.1 U/ μ l *Taq* polymerase (Applied Biosystems, Foster City, CA, USA). Reactions were cycled for 45 s at 96°C, 45 s at 68°C, and 3 min at 70°C for 24 cycles on a TGRADIENT Thermal Cycler (Biometa, Göttingen, Germany). PCR products were electrophoresed through a 20-cm 1% agarose gel (SIGMA CHEMICAL, St. Louis, MO, USA) in 0.5 \times TBE (44 mM Tris-borate, pH 8.3, 1 mM ethylenediaminetetraacetate), blotted onto Hybond Nf membranes (Amersham, Buckinghamshire, UK), hybridized with a ³²P-labelled B6.7 repeat probe and visualized by autoradiography. Allele sizes were determined by comparing size markers (1 kb plus ladder, Life Technologies,

Rockville, MD, USA). For alleles mapped by MVR-PCR, allele sizes were determined by counting repeats.

DNA sequencing

We sequenced 1 kb of DNA flanking each end of the minisatellite in eight north European individuals to search for SNPs, as well as two short alleles to confirm MVR repeat variant types. Sequencing reactions were performed using PCR-amplified double-stranded DNA according to the protocol supplied by Perkin Elmer (BigDye Terminator Cycle Sequencing reaction kit, Perkin Elmer, Foster City, CA). The sequenced products were analysed on an ABI 3100 system (Applied Biosystems).

Genotyping and haplotyping of an SNP in the 3' flanking DNA of B6.7

SNP +318A/G located 318 bp from the 3' end of the minisatellite repeat array generates a restriction fragment length polymorphism site for the restriction enzyme *Nla*III. The SNP was genotyped by PCR amplification as described earlier, with 1 μ M of primers 67BR (5'-GTCCCCTAA ACCTGGGCTCA-3') and 67F (5'-GTATTGACACGC TAGGGGTG-3'), 50–150 ng genomic DNA, and 0.1 U/ μ l *Taq* polymerase in 10- μ l reactions. Amplifications were carried out for 45 s at 96°C, 45 s at 66°C, and 3 min at 70°C for 30 cycles. PCR products (278 bp) were digested with 5 units of *Nla*III at 37°C for 1.5 h, and digested fragments resolved by electrophoresis through a 3% NuSieve gel (FMC BioProducts, Rockland, ME, USA) in 0.5 \times TBE buffer containing ethidium bromide for 3 h at 100 V. DNA fragments were visualized by ultraviolet illumination. Since the amplified flanking region has three restriction sites, apart from the polymorphic site, all PCR products were cleaved into four fragments of 121 bp, 92 bp, 47 bp, and 18 bp. The digested fragment of 121 bp in length is further cleaved into two fragments of 99 bp and 22 bp if the allele is +318A. Linkage phase between this SNP in heterozygotes and the B6.7 alleles was established by allele-specific amplification using either primer 318A (5'-GACACGCTAGGGG TGGACAT-3') or 318G (5'-GACACGCTAGGGGTG GACAC-3') together with primer 67C at the 3' end of the minisatellite, followed by analysis of minisatellite allele length.

Isolation of B6.7 alleles from genomic DNA and MVR analysis

Samples of 50–150 ng of genomic DNA were amplified as described earlier for determination of allele size and electrophoresed through a 20-cm SeaKem HGT agarose gel (FMC BioProducts) in 0.5 \times TBE buffer at 70 V for 12 h. Each allele was visualised under ultraviolet illumination after staining with ethidium bromide and the gel containing the allele was excised with a scalpel. Purified DNA was then extracted with a QiaexII Gel Extraction Kit (QIAGEN

GmbH, Hilden, Germany) and mapped by MVR-PCR as described previously (Tamaki et al. 1999).

Allele alignment

To identify alleles sharing regions of map similarity, we compared allele codes with each other by dot matrix analysis using modified Microsoft Excel software originally written by AJJ for MS32 allele analysis. Comparisons searched for perfect four-repeat matches, and allele pairs showing more than seven such matches over the best two diagonals were selected. The authenticity of these selected matches and the final alignment of allele groups were checked by eye, with gaps inserted to improve alignments.

Results

Minisatellite B6.7 allele diversity in Japanese

The size distribution of 188 Japanese alleles typed by PCR amplification is compared with north European alleles (Tamaki et al. 1999) in Fig. 1. Both populations show a unimodal size distribution, although Japanese alleles tend to be shorter than those in north Europeans (Kolmogorov-Smirnov test, $P < 0.01$) with very few large (>200-repeat) alleles. Japanese show a modal allele length of 62 repeats and a range from 8 to around 240 repeats. Seventy-four percent of these alleles are shorter than 80 repeats and can therefore be mapped completely by MVR; thus Japanese alleles are particularly suitable for structural analysis. Despite the shorter alleles, locus B6.7 remains highly polymorphic in Japanese with all of the 92 individuals typed being heterozygous for two different length alleles. We note that two of these appeared as single-band apparent homozy-

gotes by PCR typing but were shown to contain a second large (240-repeat) and poorly amplified allele by conventional Southern blot hybridisation of *Mbo*I-digested genomic DNA (Tamaki et al. 1999).

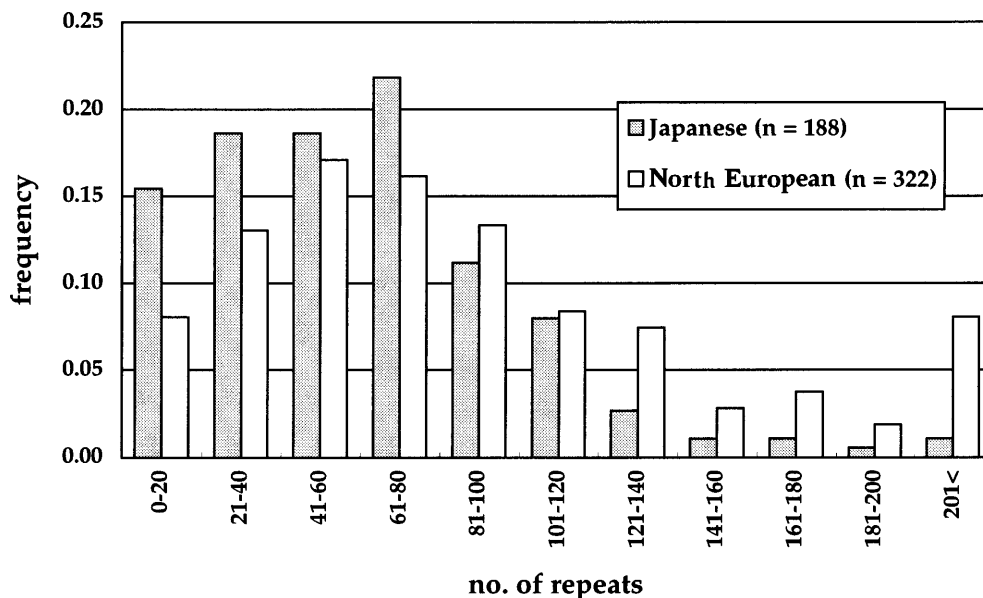
Variants flanking the minisatellite

Resequencing of 1-kb DNA flanking each end of B6.7 in the north European population revealed only one SNP. This variant, termed +318A/G, is located 318bp downstream of the repeat array, with allele G present at a frequency of 0.505 in 120 Japanese genotyped at this locus. Genotypes were in Hardy-Weinberg equilibrium ($\chi^2 = 0.034$, $df = 1$, $P = 0.853$) and the heterozygosity is high (50%).

Allele haplotyping by MVR-PCR

The paucity of SNPs near B6.7 precluded structural mapping of individual alleles from genomic DNA using allele-specific MVR-PCR directed to flanking heterozygous SNPs (Monckton et al. 1993). Instead, alleles were PCR amplified from genomic DNA and separated by gel electrophoresis prior to mapping by MVR-PCR. Examples of mapping are shown in Fig. 2. MVR codes for one set of repeat variants are read from the bottom in forward mapping, while a second set of variants are read from the top in reverse mapping. These forward and reverse codes are then combined to deduce the sequence status of each repeat unit over all typed variants (Tamaki et al. 1999); this complex multistate mapping is only possible when an allele is mapped completely in both directions. Allele B is 2027 bp in length and was mapped completely by reading the MVR status of all 58 repeats. A leading 21-bp long incomplete repeat and the first full repeat (code position 0) at the 5' end of the allele were not scored because of the weak intensity of the signal

Fig. 1. B6.7 allele size distributions in Japanese and north Europeans



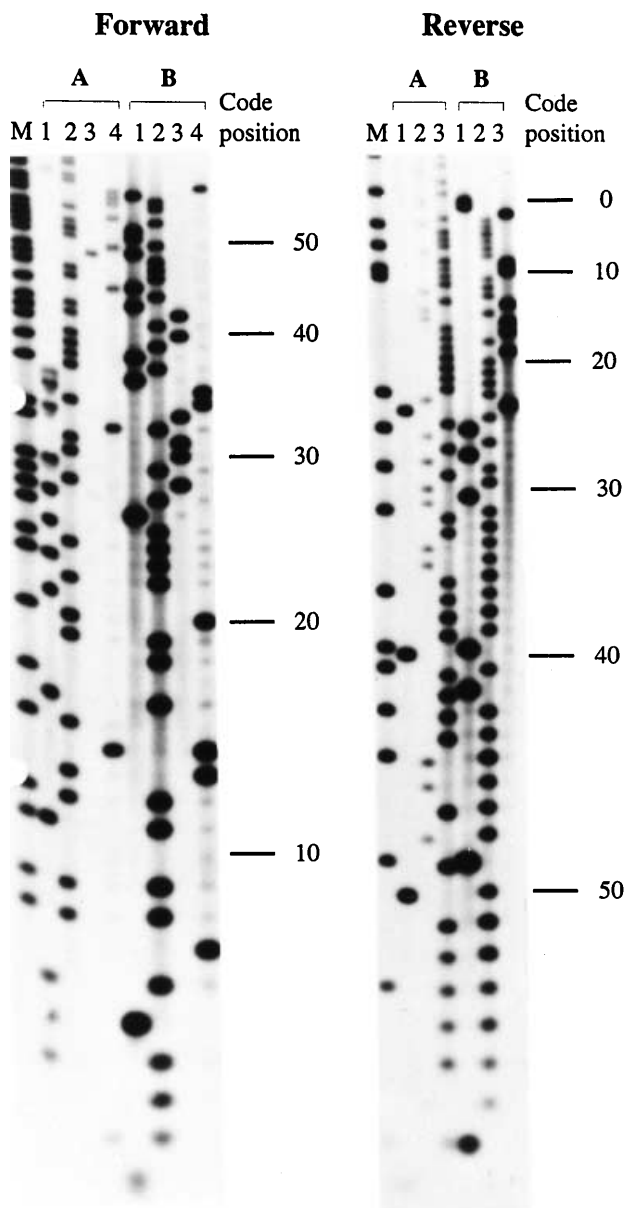


Fig. 2. Examples of forward and reverse minisatellite variant repeat mapping by polymerase chain reaction (MVR-PCR) of B6.7 alleles A and B. MVR coding positions are marked for allele B (2027 bp, 58 coded repeats plus 1.6 uncoded repeats at the 5' end). Numbers above lanes indicate the MVR-specific primers used to type polymorphic positions in each repeat as follows: forward (positions 11 and 18), 1, CG; 2, AG; 3, CA; 4, AA, and reverse (positions 5 and 7), 1, TC; 2, TT; 3, deletion and N (see also Table 1). M, Standard human DNA for determining the MVR coding position

in forward mapping. The scoring of faint MVR codes at position 1 was verified where necessary by increasing the number of PCR cycles to 27 and visualising PCR products with ethidium bromide staining. Ninety-two Japanese and 49 north European alleles were mapped by MVR-PCR.

The combined forward and reverse MVR-PCR systems can in theory distinguish 20 different MVR codes, including null repeats which contain additional variants that prevent MVR-specific primers from annealing (Tamaki et al.

1992a). All these states were detected in a total of 2281 and 2484 repeats analysed in Japanese and north Europeans, respectively. Surprisingly, the frequency of repeat types was very similar in the two populations (Table 1) despite major differences in allele lengths and structures (see following).

Identification of groups of related alleles

All 92 Japanese and 49 north European alleles mapped by MVR-PCR were compared by dot matrix analysis to identify groups of significantly related alleles. A significant match was empirically determined as at least seven four-repeat matches over the best two diagonals; variation in the match criteria had relatively little effect on the identification of related alleles (not shown). Over all 9870 possible pairwise comparisons, only 72 comparisons involving 23 north European and 35 Japanese alleles showed a match. Visual inspection of aligned alleles showed significant sharing of repeat motifs amongst groups of most of these alleles. However, alignments involving 9 of the north European alleles involved very short motifs scattered along the best diagonals and were difficult to interpret; these alleles were therefore excluded, leaving only 14 north European and 35 Japanese alleles that showed significant inter-allelic similarities, corresponding to 28% and 38% of all mapped alleles, respectively. All north European alleles were different, whereas in the Japanese only two pairs of identical alleles were found. The 12 groups of related or identical alleles identified by dot matrix analysis are shown in Fig. 3. Nine of these groups contain only Japanese alleles, and three only north European alleles. There were no instances of related alleles shared between populations.

Discussion

Minisatellite B6.7 shows huge allelic variability in Japanese as well as in north Europeans, as expected given its very high mutation rate as established in pedigrees and by direct analysis of sperm DNA (Tamaki et al. 1999). The relatively short alleles seen in Japanese make this population particularly suitable for further mutation analysis, particularly of the poorly understood processes operating in the female germline where analysis in north Europeans is limited by the fact that many mutants are too large for MVR typing. We note that MVR-PCR does not reveal all sequence variation within B6.7 alleles. To check this further, we completely sequenced three short Japanese alleles (12, 23, and 24 repeats long, data not shown). Sequences confirmed that the MVR codes are authentic and revealed additional minor sequence variations in null repeats as previously seen in north European alleles (a T→C transition at position 14 and a G→T transversion at position 21; Table 1; Tamaki et al. 1999); these additional variants should prove useful for further allele characterisation.

Almost all the Japanese alleles so far mapped by MVR-PCR are different (90 out of 92). The sampling distributions

Table 1. Proportions of repeat variants in 92 Japanese alleles and 49 north European alleles

Repeat type	Polymorphic position				Japanese No. of repeats (%)	North European No. of repeats (%)
	5	7	11	18		
a	T	C	C	G	254 (11.1)	343 (13.8)
b	T	T	C	G	123 (5.4)	233 (9.4)
c	—	N	C	G	41 (1.8)	55 (2.2)
d	T	C	A	G	23 (1.0)	40 (1.6)
e	T	T	A	G	595 (26.1)	670 (27.0)
f	—	N	A	G	409 (17.9)	317 (12.8)
g	T	C	C	A	107 (4.7)	26 (1.0)
h	T	T	C	A	17 (0.7)	50 (2.0)
i	—	N	C	A	11 (0.5)	4 (0.2)
j	T	C	A	A	3 (0.1)	21 (0.8)
k	T	T	A	A	121 (5.3)	148 (6.0)
l	—	N	A	A	72 (3.2)	68 (2.7)
m	T	C	u	u	42 (1.8)	60 (2.4)
n	T	T	u	u	226 (9.9)	278 (11.2)
p	—	N	u	u	178 (7.8)	71 (2.9)
q	u	u	C	G	14 (0.6)	20 (0.8)
r	u	u	A	G	26 (1.1)	45 (1.8)
s	u	u	C	A	3 (0.1)	3 (0.1)
t	u	u	A	A	5 (0.2)	6 (0.2)
o	u	u	u	u	11 (0.5)	26 (1.0)
Total					2281 (100.0)	2484 (100.0)

B6.7 repeat sequence: 5'-TCTCTAYAGG**MC**AYGAG**RG**TKGACAGTGAGGGG-3' (the four polymorphic positions are shown in bold)

—, Deletion of T; u, detected as null repeats in forward and/or reverse minisatellite variant repeat mapping due to other base variants that block annealing by MVR-specific primers (Tamaki et al. 1992a)

of different alleles can be used to estimate allele diversity, $\theta = 4N_e\mu$, where N_e is the effective population size and μ is the mutation rate. Under the infinite allele model and assuming selective neutrality, the number of different alleles, n_a , seen in a sample of i individuals is given by $n_a = \sum_i \{i/(i+1)\}$ (Ewens 1972) and heterozygosity can be estimated as $\theta/(1+\theta)$. The Japanese data give an estimated θ value of 2030 for the Japanese and, assuming a value of 10,000 for N_e (Neel et al. 1978), implies a mutation rate of 5% per gamete, very similar to the sex-averaged mutation rate at B6.7 seen in pedigrees (Olaisen et al. 1993). This estimate of diversity also suggests a heterozygosity of 99.95%. If all alleles are equally rare, Poisson analysis indicates that more than 2000 different Japanese B6.7 alleles must exist to result in this sampling frequency distribution. Since allele frequencies will not be uniform, the true level of allele diversity is likely to be much higher.

Despite this huge diversity, groups of structurally related alleles can be identified, each of which is presumably derived by mutational divergence from a common ancestral allele. Remarkably, these groups appear to be completely specific to Japanese or to north Europeans, presumably as the result of rapid mutation after the recent divergence of these two populations. Consistent with this, analysis of more stable minisatellites such as MS31, MS32, and MS205 has shown less extreme population specificity, with groups of related alleles often only partially specific to a given population and frequently shared between populations (Jeffreys et al. 1991; Tamaki et al. 1992b; Neil and Jeffreys 1993; Huang et al. 1996; Armour et al. 1996). Allele groups identified at highly unstable loci such as B6.7 may therefore

provide useful markers for exploring very recent events such as population bottlenecks, expansions, and migration.

Differences between groups of closely related B6.7 alleles (Fig. 3) appear to be randomly scattered along the repeat array, consistent with the lack of mutational polarity (mutation events restricted to one end of the array) seen in sperm mutants (Tamaki et al. 1999). There is, however, a tendency for the ends of the array to be relatively invariant within some groups (groups I-V, VII, VIII, XI), suggesting that the mutation process to some extent avoids these termini. There are very few pairs of alleles that can be related by a simple mutation event (e.g., the very short alleles 9 and 10 that are related by gain/loss of a single e-type repeat unit). Again, this is consistent with the very complex recombinational rearrangements frequently seen in sperm mutants (Tamaki et al. 1999). Allele groups also show evidence for large duplications of blocks of repeats (e.g., allele 27) and single repeat switches without change in repeat copy number (e.g., the beginnings of alleles 19–21); similar rearrangements have also been directly observed in sperm. Thus, allele diversity mirrors the mutants seen in sperm, although it cannot be used to determine mutation processes given the impossibility of identifying pairs of alleles separated by a single mutation event (e.g., alleles 16 and 17 may have diverged by a series of separate relatively small events scattered along the array, or instead by a single but very complex mutation event).

The almost complete absence of SNPs around B6.7, which stands in contrast to the abundance of SNPs seen near other minisatellites (Jeffreys et al. 1998b; Buard et al. 2000; Armour et al. 1993), has prevented us from placing

Fig. 3a,b. Groups of aligned B6.7 alleles identified initially by dot matrix analysis. **a** Japanese alleles, **b** north European alleles. The number of repeats including the uncoded first repeat at the 5' end is given for each allele, together with MVR structure and 3' flanking haplotype at marker +318A/G. The flanking haplotype in north Europeans was not determined. The main MVR segments shared by related alleles are shown in *red*, and other shared segments restricted to subgroups of alleles in *additional colors*. Gaps have been introduced to improve alignments. Intraallelic duplications are indicated by *double underlining* of the 5' duplicated unit and *single underlining* of the 3' unit. Alleles marked with *asterisks* were observed twice in the present study

		a	
		Japanese alleles	MVR haplotype
No.	length		+318
I			
1	42	aamknene f e f e f e e n n a p n g f b p g f b f e g f f e b g f e e b p g	A
2	36	<u>a f e e e</u> -- b p g f b f e g f f e e e e f e - g f f e e e p p - n k g	A
3	25	g e f a t b k f m t b f f e e e a e e p p - n k g	A
4	36	a f e e e p e p f n e n p p n p k e e e e f e - - k e e e p p - n k g	A
5	38	a f e e e e e f e n f f f f f p r e e e e - f e e n e f f e n p p - n k g	A
6	31	a f e e e e e f e p f f f f f - - - - - e e f a n n n n p p n k g	A
7	12	a f e e e e - - - - - p p - h k g	A
8	11	a f e e e - - - - - p p - k k g	A
9	11	a f e e e - - - - - p p - n k g	A
10	10	a f e e - - - - - p p - n k g	A
11	18	a f e e b l e e e - - - - - b a e p p - n e g	A
12	26	a f e e b l e e e - - - - - n e e e n r p n e a e p - n e g	A
13	18	a f e e - - - - - e e e n e n p a e p - n e g	A
14	19	a f e e - - - - - k e e n e p p a e p - n e g	A
15	27	a f e - - - - - n f e e n e e e e n n e n n a e p - n e g	A
II			
16	58	m e f e b f l e - - a e b f b f e - - - - a l f e a a l f a a f b f e a n r - - - a f f f e a f a e a l b - f e f f b f e e a f a	A
17	63	<u>g e f e b f l e</u> - - a e b f b f e a f f e a l f e a a l f a a f b f e a n g f a a e a f f e - - a a e a - - f f - f f b f e e a f a e a	G
18	47	<u>i e f e b f l e l k</u> a e b f b f e - - - f a l f e a a - - - - - f f - a f a e a - - f f e f f b f e e a f a e a	G
III			
19	31	a f e e e e e f e p f f f f f - - - - - e e f a n n n n p p n k g	A
20	38	a f e e e e e f e n f f f f f p r - - - - - e e e e e e n e f f e n p p n - k g	A
21	46	a f e e e e e f e p r f f f f <u>a f e e e f a e e a f e a g n e f f e n p p n e a e g</u>	A
IV			
22	41	a a m k k n k n p a e p a f e n f b f c h p m c e e c e h - - - p p p n c p f e b p g	A
23	42	a a m k k n k n p a e p a f e n f b f c h p n c - - f e h f f c n p p n c p p n b p g	A
24	36	a a m k k n k n p a e p a f e n f b f c h p n c e e f e c f f - - - - - b b p j	A
V			
25	23	a c e e a f f b e e n o - - - - - e p b e a e p n e g	A
26	35	a f e e a f f e e p p n e b e e e n r e b e p p e p b e a e p n e g	A
VI			
27	38	a a m k k n n n m a k n e e e e l k f f k n n p m a k n e e e l k f k i	G
28	21	a a m k k n n p a a k n e e e f n f l g	A
VII			
*29	14	a a m k - - - - - e n g f f f b p g	A
30	19	a q m k f f e f f e n g f f b p g	A
VIII			
31	11	a k m k n - - - - - k k f l g	G
32	24	a k m k n e e a n n p k k k n k n k k f l g	G
IX			
*33	19	g e f f b p r c e a f f e a f a e a	G
b			
		North European alleles	MVR haplotype
No.	length		
X			
34	67	a a e m a r e a a r b a j e k b e j h h d e e k e k e j e e f j e e d b e d b e d f k d e n g e a n g e d e k e d e f e f n e	
35	84	a a e m a r e a a r b a a r e a p a e n l f e f b d e t f f m f c e e n m f f e k c f f e k e m f f f k c f f m e f r f e b f r n e f e c b e b f c e f m e a e f e	
36	39	a a e m a r e a a r b a a r e a p a e c l e k n a m k n a a i l e l k n r g	
37	69	a a e m a r e a a r b a a r e - p a e t f f r f f e d p e f a n b f c a c a e d a b h c - - - - - h l f l f f e k k k f e e n f a e a h k p ? ?	
38	64	n b e f d f f a n e f a n e f c a c a e d a b h c f e e k f e e k k f e e k f e e k k f a e k k f k f e n n f a e a h k d	
39	45	a a e m a r e a a r b a a r e - p a e t f f r f f e - - - - - f a e f f n h e n f a e e a f g k d	
40	48	a a - - - r e a a r b a a r e - p a e t f f e d f f e - - - - - a e p e f e f e f c a c o f e b f a e a h k d	
XI			
41	22	m a - n n e - - - - - c a n n e n r n - - - - - f p r n h n h e	
42	40	m a - n n e - - - - - c a n n - - - - - n f b n e b - - - n r n e e p r e n b h n h f n n n r n h n h - n e	
43	51	m a - n n e - - - - - c a n n - - - b b k f n g f n e m n e n - - - b b n e f n e m n e e e a e e a e a p r n h n h - n e	
44	67	? a e n n e c b n n e n j n e e k e e f c a n n e n r n e p r n f n f n e m n e - n e n f - n e b b e a b n e a m n e - - - - - p r n h n h n n e	
45	66	m a e n n e c b n n e n j n e e k e e f c a n n e n r n e p r n f n f n e m n e - n e n f - n e b - e a b n e a m n e - - - - - p r n h n h n n e	
XII			
46	86	a f a e a l p n k e e e a m o n e a a e f p e e e f f n f e d j e j f e k j g e j e j b l e a b k e o k f e e j e j f e k j g e j e j b l e a b k e k e a a f a a n a	
47	57	a f a e a l p ? ? a e e m n n e e e e e a e b e b c k e e b n r p r f a e e e f f d k f e k e j b a a	

B6.7 diversity within the context of an extended flanking haplotype (Armour et al. 1996). In most groups of aligned alleles, all B6.7 alleles share the same haplotype at the solitary 3' marker +318A/G. There are however two exceptions, in groups II and VI (Fig. 3), both of which show switches at this SNP plus reorganisation of the 3' end of the array. This suggests that B6.7 may also be proficient in meiotic crossover, leading to exchange of this flanking marker, and/or in gene conversion events that extend into flanking DNA. Both types of events have been observed in sperm mutants at minisatellites MS32 and CEB1 and at MS32 define a flanking recombination hot spot that is apparently responsible for driving repeat instability (Jeffreys et al. 1998a,b; Buard et al. 2000). The lack of SNPs near B6.7 will make direct analysis of such recombination events very difficult, all the more frustrating given that this subtelomeric minisatellite is located in a region remarkably rich in other minisatellites, with repeat arrays located 1 kb 5' and 0.3 kb 3' of the B6.7 minisatellite (Tamaki et al. 1999 and unpublished data).

Acknowledgments We are grateful to Professor Mitsuru Mori for statistical advice and Charles Brenner for helpful comments. This work was supported by grants from the Minister of Education, Science, Sport and Culture of Japan.

References

- Andreassen R, Olaisen B (1998) De novo mutations and allelic diversity at minisatellite locus D7S22 investigated by allele-specific four-state MVR-PCR analysis. *Hum Mol Genet* 7:2113–2120
- Armour JAL, Harris PC, Jeffreys AJ (1993) Allelic diversity at minisatellite MS205 (D16S309): evidence for polarized variability. *Hum Mol Genet* 2:1137–1145
- Armour JAL, Anttinen T, May CA, Vega EE, Sajantila A, Kidd JR, Kidd KK, Bertranpetit J, Pääbo S, Jeffreys AJ (1996) Minisatellite diversity supports a recent African origin for modern humans. *Nat Genet* 13:154–160
- Buard J, Vergnaud G (1994) Complex recombination events at the hypervariable minisatellite CEB1 (D2S90). *EMBO J* 13:3203–3210
- Buard J, Bourdet A, Yardley J, Dubrova Y, Jeffreys AJ (1998) Influences of array size and homogeneity on minisatellite mutation. *EMBO J* 17:3495–3502
- Buard J, Shone AC, Jeffreys AJ (2000) Meiotic recombination and flanking marker exchange at the highly unstable human minisatellite CEB1 (D2S90). *Am J Hum Genet* 67:333–344
- Ewens WJ (1972) The sampling theory of selectively neutral alleles. *Theor Popul Biol* 3:87–112
- Huang X-L, Tamaki K, Yamamoto T, Suzuki K, Nozawa H, Uchihi R, Katsumata Y, Neil DL (1996) Analysis of allelic structures at the D7S21 (MS31A) locus in the Japanese, using minisatellite variant repeat mapping by PCR (MVR-PCR). *Ann Hum Genet* 60:271–279
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Jeffreys AJ, Neumann R, 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
- Jeffreys AJ, MacLeod A, Tamaki K, Neil DL, Monckton DG (1991) Minisatellite repeat coding as a digital approach to DNA typing. *Nature* 354:204–209
- Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nat Genet* 6:136–145
- Jeffreys AJ, Murray J, Neumann R (1998a) High resolution mapping of cross-overs in human sperm defines a minisatellite-associated recombination hotspot. *Mol Cell* 2:267–273
- Jeffreys AJ, Neil DL, Neumann R (1998b) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J* 17:4147–4157
- Kimpton CP, Hopgood R, Watson SK, Gill P, Sullivan K (1992) Cloning and characterisation of novel single locus probes for forensic purposes. In: Rittner C, Schneider PM (eds) *Advances in forensic haemogenetics* 4. Springer-Verlag, Berlin Heidelberg New York, pp 129–131
- Monckton DG, Tamaki K, MacLeod A, Neil DL, Jeffreys AJ (1993) Allele-specific MVR-PCR analysis at minisatellite D1S8. *Hum Mol Genet* 2:513–519
- Neel JV, Ueda N, Satoh C, Ferrell RE, Tanis RJ, Hamilton HB (1978) The frequency in Japanese of genetic variants of 22 proteins. V. Summary and comparison with data on Caucasians from the British Isles. *Ann Hum Genet* 41:429–441
- Neil DL, Jeffreys AJ (1993) Digital DNA typing at a second hypervariable locus by minisatellite variant repeat mapping. *Hum Mol Genet* 2:1129–1135
- Olaisen B, Bekkemoen M, Hoff-Olsen P, Gill P (1993) Human VNTR mutation and sex. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds) *DNA fingerprinting, state of the science*. Birkhauser Verlag, Basel, pp 63–69
- Royle NJ, Clarkson RE, Wong Z, Jeffreys AJ (1988) Clustering of hypervariable minisatellites in the proterminal regions of human autosomes. *Genomics* 3:352–360
- Tamaki K, Monckton DG, MacLeod A, Neil DL, Allen M, Jeffreys AJ (1992a) Minisatellite variant repeat (MVR) mapping: analysis of “null” repeat units at D1S8. *Hum Mol Genet* 1:401–406 [Erratum (1992) *Hum Mol Genet* 1:558]
- Tamaki K, Yamamoto T, Uchihi R, Kojima T, Katsumata Y, Jeffreys AJ (1992b) DNA typing and analysis of the D1S8 (MS32) allele in the Japanese population by the minisatellite variant repeat (MVR) mapping by polymerase chain reaction (PCR) assay. *Jpn J Legal Med* 46:474–482
- Tamaki K, May CA, Dubrova YE, Jeffreys AJ (1999) Extremely complex repeat shuffling during germline mutation at human minisatellite B6.7. *Hum Mol Genet* 8:879–888