J Mol Evol (1997) 44:336–340

JOURNAL OF MOLECULAR EVOLUTION

Microsatellite Allelic Homoplasy Due to Variable Flanking Sequences

Marie-Claude Grimaldi, Brigitte Crouau-Roy

Centre d' Immunopathologie et de Génétique Humaine, CNRS, UPR 8291, CHU Purpan, 31300 Toulouse, France

Received: 14 May 1996 / Accepted: 9 September 1996

Abstract. Microsatellite DNA sequences have become the dominant source of nuclear genetic markers for most applications. It is important to investigate the basis of variation between alleles and to know if current assumptions about the mechanisms of microsatellite mutation (that is to say, variations involving simple changes in the number of repeat) are correct. We have characterized, by DNA sequencing, the human alleles of a new highly informative (CA)n repeat localized approximately 20 kb centromeric to the HLA-B gene. Although 12 alleles were identified based on conventional length criteria, sequencing of the alleles demonstrated that differences between alleles were found to be more complex than previously assumed: A high degree of microsatellite variability is due to variation in the region immediately flanking the repeat. These data indicate that the mutational process which generates polymorphism in this region has involved not only simple changes in the number of dinucleotide CA repeats but also perturbations in the nonrepeated 5′ and 3′ flanking sequences. Three families of alleles (not visible from the overall length of the alleles), with presumably separate evolutionary histories, exist and can yield to homoplasy of size. Effectively, we can observe alleles of the same size with different internal structures which are separated by a significant amount of variation. Although allelic homoplasy for noninterrupted microsatellite loci has been suggested between different species, it has not been unequivocally demonstrated within species. A strong association is noted between alleles defined at the sequence level and HLA-B alleles. The observation of several families of alleles at the population level provides information about the evolutionary history and mutation processes of microsatellites and may have implications for the use of these markers in phylogenetic, linkage disequilibrium studies, and gene mapping.

Key words: Microsatellite — Homoplasy — Evolution — Sequence

Introduction

Microsatellite loci are short tandem arrays of DNA widely dispersed throughout the nuclear genome in eucaryotes. They are often highly polymorphic due to variation in the number of repeat units (Weber and May 1989; Weber 1990; Kashi et al. 1990; Henderson and Petes 1992). Microsatellites provide a major source of molecular markers for population genetics (Deka et al. 1991, 1995; Crouau-Roy et al. 1993; Slatkin 1995), evolutionary studies (Bowcock et al. 1994; Goldstein et al. 1995), linkage analysis (Todd et al. 1991; Dietrich et al. 1992), and human gene mapping (Weissenbach et al. 1992).

However, the precise nature of mutations of microsatellite sequences and how they evolve in natural populations are far from clear. It seems likely that slippage during replication is one of the causes of the observed length polymorphism (Strand et al. 1993; Weber and Wong 1993). New experimental and analytical data (Schlötterer and Tautz 1992; Henderson and Petes 1992; Sirugo et al. 1992; Oudet et al. 1993; Valdes et al. 1993) suggest that the population genetic process able to account for the distribution of allele frequencies is loss or

Correspondence to: B. Crouau-Roy; e-mail bcrouau@cict.fr

gain of generally one repeat unit at a time (consistent with a ''stepwise mutation'' model). Di Rienzo et al. (1994) introduced a two-phase mutation model, in which most mutations are single-step changes, but rare large jumps in repeat number also occur, a model supported by empirical evidence (Farber et al. 1994) and by direct analysis of new mutant alleles in pedigrees (Chung et al. 1993; Zhong et al. 1993; Shibata et al. 1994; Wooster et al. 1994).

It is important that the assumption of microsatellite variability due to simple repeat variation be tested because microsatellites have become the markers of choice for population genetic studies and are highly informative markers for multigenic and multifactorial diseases and human linkage studies. Garza et al. (1995) and Blanquer-Maumont and Crouau-Roy (1995) showed that differences between alleles of different species are not always attributable to change in the number of repeat units but also may be due to changes in the composition of the repeat element or to insertions or deletions of sequence flanking the repeats. At the intraspecific level, there is a lack of data relating to microsatellite structure; the previous studies concern allele structure of interrupted or compound microsatellites (Adams et al. 1993; Urquhart et al. 1993; Estoup et al. 1995). We identified a new (CA)n repeat from the human major histocompatibility complex region (MHC) near the HLA-B gene and analyzed its polymorphism and the association with HLA-B in a sample consisting of unrelated individuals already typed for the HLA-B gene and in reference panels of HLA-typed cell lines (Abraham et al. 1993). Assumptions about the mechanisms of microsatellite mutation were tested by direct sequencing of different alleles, and particularly alleles of identical size but associated with different HLA-B alleles.

Materials and Methods

The microsatellite, called MIB, was identified in the clone 17.1, described by Geraghty et al. (1992), which includes the pseudogene HLA 17. Its position, established by DNA sequence analysis, is about 20 kb centromeric to HLA-B.

DNA Analysis. We analyzed a total of 110 DNA samples from HLA-typed individuals distributed as follow: 45 unrelated French Caucasoids, 43 cell lines from the cell panel of the 4 Asia Oceania Histocompatibility Workshop (4AOH, 1993) (some of which were studied in the 10th International Histocompatibility Workshop, 10IHW), and two families (a nuclear family of ten individuals from European Collection Biomedical Rechearch [ECBR, 1992] and a three-generation family of 12 individuals from CEPH [Centre d'Etude du Polymorphisme Humain]). DNA was isolated from lymphoblastoid cell lines or peripheral blood leucocytes.

Repeat-flanking oligonucleotide primers were synthesized as follows: MIB(A): 5′CTACCATGACCCCCTTCCCC 3′ MIB(B): 5′CCACAGTCTCTATCAGTCCA 3′.

The PCR reactions were performed using end-labeled oligonucleotide primers in a final reaction volume of 25 μl with 50 ng genomic DNA, 4 μl dNTP (1.25 mM), 1.5 mM $MgCl_2$ (final concentration), 25 pmol cold forward primer, 17.5 pmol cold reverse primer plus 0.04 pmol end-labeled reverse primer $\gamma[^{32}P]$ATP, 0.5 U Taq polymerase (Promega) in a total volume of 25 μl. After initial denaturation for 5 min at 95°C, amplifications were carried out in an automated DNA thermal cycler (Perkin Elmer 480) for 30 cycles consisting of 1 min at 95°C, 1 min at 55°C, and 1 min at 72°C with a final extension of 4 min at 72°C. A 5-μl aliquot of each reaction was analyzed on a 6% denaturing polyacrylamide (sequencing) gel and alleles were visualized by autoradiography on XAR (Kodak). A sequencing ladder was used to determine size differences between alleles.

Microsatellite Sequencing. The DNA sequences of different microsatellite alleles were determined. Fragments were amplified from genomic DNA by PCR in nonradioactive 50-μl reactions (conditions as above). The amplified products were ligated into a TA cloning vector (Invitrogen) and cloned into DH5 α-competent cells. Plasmid DNA was extracted and purified using Wizard Minipreps (Promega). The nucleotide sequencing of double-stranded DNA was carried out by the dideoxytermination method (Sanger et al. 1977) using T7 DNA polymerase (Pharmacia). We sequenced several clones for each French individual to eliminate the errors resulting from Taq polymerase misincorporation or in vitro recombinant PCR products (Ennis et al. 1990). Only those clones which yielded amplification products identical in size to that obtained using genomic DNA from the same individual were sequenced. Moreover, for alleles 1 and 9, sequences from different individuals have been obtained.

Results

Twelve alleles were identified with sizes between 326 bp (allele MIB1) and 356 bp (allele MIB12). The analysis of DNA samples from the unrelated French sample revealed a heterozygosity of 0.85 and a *PIC* (Polymorphic Information Content) value 0.78, reflecting the great informativity of this microsatellite. The estimated allele frequencies and their support intervals are displayed in Table 1. All associations observed in the homozygous cells from the HLA Workshop were compatible with the French defined haplotypes (data not shown). A good correlation is observed between allelic variation at the HLA-B locus and the microsatellite polymorphism: For example, the HLA-B 62 allele is always associated with MIB 1, HLA-B 14 with MIB 2, and HLA-B 35 and 60 with MIB 9.

All the MIB alleles differ by multiples of two base pairs, so it is easy to assume that allelic differences are entirely due to changes in the number of the basic repeat unit, but there was no direct evidence for this. All sizes of allele (and associated with several HLA-B alleles) were sequenced except the rare allele MIB5: From the heterozygote MIB5 individuals, only the non-MIB5 alleles were cloned and no cell lines were found homozygous with it. The sequences containing the (CA)n repeats and their immediate 3′ and 5′ flanking regions are shown in Table 1. Outside this region, the sequence is, for all the alleles, identical to that published in Geraghty et al. 1992.

Variation at this locus is attributable to differences in both the number of CA repeats and the directly flanking regions. Three types of polymorphism are observed in

Table 1. Allele frequencies associated with their support intervals and sequences of the MIB microsatellite alleles (classified by the length of the PCR products) from unrelated French Caucasian individuals[a]

MIB alleles	Allele frequencies	Number of bases	Sequences
1	0.266 (0.132–0.320)	326	. . . GTTATCC**C**C**(CA)$_7$** TAAA**C**ATAAACATAAGTAT. .
2	0.056 (0.046–0.070)	332	. . . GTTATCC- - **(CA)$_{11}$**TAAA**C**ATAAACATAAGTAT. .
3	0.289 (0.209–0.302)	334	. . . GTTATCC- - **(CA)$_{12}$**TAAA**C**ATAAACATAAGTAT. .
4	0.044 (0.017–0.070)	338	. . . GTTATCC- - **(CA)$_{14}$**TAAA**C**ATAAACATAAGTAT. .
6	0.044 (0.017–0.070)	342	. . . GTTATCC- - **(CA)$_{19}$**TAAA- - - - - - CATAAGTAT. .
7	0.011 (0.006–0.024)	346	. . . GTTATCC- - **(CA)$_{21}$**TAAA- - - - - - CATAAGTAT. .
8	0.100 (0.051–0.231)	348	. . . GTTATCC- - **(CA)$_{19}$**TAAA**C**ATAAACATAAGTAT. .
9	0.100	**350**	. . . GTTATCC- - **(CA)$_{23}$**TAAA- - - - - - CATAAGTAT. .
	(0.051–231)	**350**	. . . GTTATCC- - **(CA)$_{20}$**TAAA**C**ATAAACATAAGTAT. .
10	0.033 (0.014–0.054)	352	. . . GTTATCC- - **(CA)$_{21}$**TAAA**C**ATAAACATAAGTAT. .
11	0.022 (0.011–0.031)	354	. . . GTTATCC- - **(CA)$_{25}$**TAAA- - - - - - CATAAGTAT. .
12	0.022 (0.010–0.033)	356	. . . GTTATCC- - **(CA)$_{23}$**TAAA**C**ATAAACATAAGTAT. .

[a] Allele 5 not sequenced (see text)
Observed variations are in boldface

the PCR product. The alleles (defined by the size of PCR product) differ mainly in variation in the number of (CA) repeats from seven CA (allele 1) to 25 CA repeats (allele 11). However, they also differ in the presence or absence of a duplication of the nucleotides (CATAAA) in the sequence 3′ of the repeat: This duplication is observed in eight different alleles. Thus, for example, the difference between alleles 7 and 8, which differ in size by two base pairs, is not due to the addition of one repeat unit but to a decrease of two CA repeats along with duplication of the motif CATAAA. Moreover, in the 5′ part of the sequence, there were two nucleotides C in all alleles except for MIB1, where there were four repetitions of the cytosine.

The microsatellite alleles associated with the HLA-B alleles 35 and 60 are identical in size (MIB9 with 350 bp) and thus indistinguishable by conventional microsatellite PCR typing, but their sequences are strikingly different (Table 1). Their identical size is merely coincidence: The allele associated with HLA-B35 has 20 (CA) repeats and a duplication of the sequence CATAAA (350 bp), while its homologue has 23 (CA) repeats (350 bp). There was an individual heterozygous for HLA-B (with the two alleles B35 and B60; cell 4AOH No 100013G) who was homozygous for the MIB locus (allele 9), but with both sequences for this allele.

Discussion

Accumulation of DNA sequence data from the different HLA genes has allowed study of allelic variation and possible mutational processes, but for microsatellite loci there remains a shortage of data on the sequences of the different alleles within the same species. Understanding the evolution of such repetitive sequences is important as it is tied to the mechanism of mutation and amplification, which generates the high polymorphism. The fact that changes in allele size, for dinucleotide microsatellites, generally involves an increase or decrease of two pairs, or a multiple of two pairs, strongly suggested that all or nearly all of the observed changes were within the tandem repeats rather than within the nonrepeated flanking DNA (Sirugo et al. 1992; Kwiatoski et al. 1992; Oudet et al. 1993; Weber et Wong 1993). However, there are several studies providing empirical evidence for microsatellite complexity. Studies of microsatellites in the TNF region (in MHC of human and primates) and in chromosome 4 indicate that interspecific differences in allele length are not always due to simple changes in the number of repeats but also to insertion/deletion events in the sequences that flank the repeat and interruptions of perfect repeats which are correlated with stability of microsatellite (Blanquer-Maumont and Crouau-Roy 1995; Crouau-Roy et al. 1996). Garza et al. (1995) obtained similar results, finding differences in the composition of the repeat element between human and chimpanzee. Deka et al. (1995) found that for two markers the allele sizes are not always in increments of 2 bp and suggest that allele size alterations may involve insertion/deletion of single nucleotides or other, more complex phenomena.

At the population level, sequence data on noninterrupted microsatellites are not available to demonstrate the microsatellite complexity and thus a homoplasy of size. In the MIB microsatellite, direct sequencing of the different alleles observed, defined by length, reveals that the alleles present differ not only in the number of CA repeat units but also in the structure of the immediate flanking regions. Two other forms of variation are present: the presence or absence of a duplication of (CATAAA) 3′ of the CA repeat, and for one allele only (MIB1), four cytosines (produced either by duplication of the cytosine pair or by transversion A to C in the first repeat). DNAs from several individuals carrying the MIB1 allele (HLA homozygous cell lines from the

4AOH) have been sequenced, and confirm that this allele is always and only produced by seven (CA) repeats and duplication of the motif CATAAA. Thus, allelic differences are not always due to changes in the number of dinucleotide repeats. For MIB, the three families of alleles observed with presumably separate evolutionary histories can lead to a homoplasy of size. We can see this phenomena for one allele (MIB9), associated with two different alleles of HLA-B, which displays different sequences of the same length (250 bp). These sequences are consequently separated by a significant amount of variation. For the allele MIB5 (340 bp), not sequenced, we can hypothesize either 15 CA repeats with the duplication of the motif CATAAA, or 18 CA repeats. The observation of such families of alleles suggests that when multiple MIB alleles are associated with a single HLA-B allele, these forms should all belong to the same family. This remains to be demonstrated.

In microsatellites, arrays increase or decrease, by replication slippage for most of them, of one or two repeats, with occasional larger step changes (Edwards et al. 1992; Valdes et al. 1993; Di Rienzo et al. 1994). Strand et al. (1993) and Lindahl (1994) showed that there is apparently a relation between replication slippage of microsatellite sequences and defective DNA repair. However, the larger increases associated with certain trinucleotide repeats, which cause several human genetic diseases, seem to involve other, unknown, mechanisms (Kuhl and Caskey 1993; Fu et al. 1991; Zhong et al. 1993). The mechanisms of mutation of the variations, observed at the population level in the MIB microsatellite, are probably of several kinds. It is likely that there is strand slippage for the (CA)n repeat sequence; however, another mutational process seems involved in the nonrepeated sequence flanking the CA. Several works suggest that sequences flanking microsatellites could be relatively unstable, instability either caused by the CA repeat region itself or only associated with it. For example, a recent study of the regions flanking (CT)n repeats in mammalian genomes revealed that a disruption in homology between species occurred, often on the immediate 3′ or 5′ side of the repeat array (Stallings 1995). The insertion/deletion events, observed in the TNF microsatellites, also always involved the immediate nonrepeated flanking sequence (Blanquer-Maumont and Crouau-Roy 1995). Studies of Callen et al. (1993) and Koorey et al. (1993) show examples of ‘‘null’’ alleles in humans in which primers flanking microsatellite arrays fail to bind because of modifications occurring within the priming sequence for PCR amplification of the (CA)n repeats.

The observation of a substantial proportion of microsatellite variability due to sequence variation in the region flanking the repeat and consequently of some cases of homoplasy may have several implications. Sequence analysis of a large number of loci within and between species might be useful to elucidate the evolutionary history, mutation processes, and role in the genome evolution of these repeat sequences. This can also improve mapping for genetically complex traits, allowing one to distinguish microsatellite alleles based on their ancestral relationship. Finally, this would provide more information to assess the existence of linkage disequilibrium between linked markers.

Acknowledgments. We wish to thank N. Freimer and his group for helpful discussions and the reading of the manuscript. Thanks are also due to J. Clayton for English corrections, to A. Cambon-Thomsen for the cell lines, to P. Pontarroti and R. Jambou for help with cell culture, and to C. Bouissou for technical assistance. This work was supported by the French Ministere de la Recherche, the National Center of the Scientific Research (CNRS), and the European Community.

References

Abraham L, Marley J, Nedospasov S, Cambon-Thomsen A, Crouau-Roy B, Dawkins R, Giphart M (1993) Microsatellite, restriction fragment-length polymorphism, and sequence-specific oligonucleotide typing of the tumor necrosis factor region. Comparisons of the 4AOHW cell panel. Hum Immunol 38:17–23

Adams M, Urquhart A, Kimpton C, Gill P (1993) The human D11S554 locus: four distinct families of repeat pattern alleles at one locus. Hum Mol Genet 2 (9):1373–1376

Blanquer-Maumont A, Crouau-Roy B (1995) Polymorphism, monomorphism and sequence in conserved microsatellites in Primates species. J Mol Evol 41:492–497

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Munch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Callen DF, Thompson AD, Shenm Y, Phillipsm HA, Richards RI, Mulley JC, Sutherland GR (1993) Incidence and origin of ‘‘null’’ alleles in the (AC)n microsatellite markers. Am J Hum Genet 52: 922–927

Chung MY, Ranum MPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 5:254–258

Crouau-Roy B, Briant L, Pociot F, Stravopoulos A, Cambon-Thomsen A, Clayton J (1993) Tumor necrosis factor microsatellites in four european populations. Hum Immunol 38:213–216

Crouau-Roy B, Service S, Slatkin M, Freimer N (1996) A fine scale comparison of the human and chimpanzee genomes: linkage, linkage disequilibrium, and sequence. Hum Mol Genet 5:1131–1137

Deka R, Chakraborty R, Ferrell RE (1991) A population genetic study of six VNTR loci in three ethnically defined populations. Genomics 11:226–230

Deka R, Jin L, Shriver MD, Yu LM, DeCroo S, Hundrieser J, Bunker CH, Ferrell RE, Chakraborty R (1995) Population genetics of dinucleotide (dC-dA)n.(dC-dT)n polymorphisms in world populations. Am J Hum Genet 56:461–474

Dietrich W, Katz H, Lincoln SE, Shin HS, Friedman J (1992) A genetic map of the mouse suitable for intraspecific crosses. Genetics 131: 423–447

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M, Freimer NB (1994) Mutational processes of simple sequence repeat loci in human populations. Proc Natl Acad Sci USA 91:3166–3170

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992) Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. Genomics 12:241–253

Ennis PD, Zemmour J, Salter RD, Parham P (1990) Rapid cloning of HLA-A, B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc Natl Acad Sci USA 87:2833–2837

Estoup A, Tailliez C, Cornuet J-M, Solignac M (1995) Size homoplasy and mutational processes of interrupted microsatellites in two bee species, Apis mellifera and Bombus terrestris (Apidae). Mol Biol Evol 12(6):1074–1084

Farber RA, Petes TD, Dominska M, Hudgens SS, Liskay RM (1994) Instability of simple sequence repeats in a mammalian cell line. Hum Mol Genet 3:253–256

Fu YH, Kuhl DP, Pizzuti A, Pieretti M, Sutcliffe JS, Richards S, Verkerk AJ, Hloden JJ, Fenwicck RG, Warren ST (1991) Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047–1058

Garza JC, Slatkin M, Freimer NB (1995) Microsatellite allele frequencies in humans and chimpanzees with implications for constraints on allele size. Mol Biol Evol 12:594–603

Geraghty DE, Koller BH, Hansen JH, Ork HT (1992) The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. J Immunol 149:1934–1946

Goldstein DB, Linares AR, Feldman MW, Cavalli-Sforza LL (1995) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471

Henderson ST, Petes TD (1992) Instability of simple sequence DNA in Saccharomyces cerevisiae. Mol Cell Biol 12:2749–2757

Kashi Y, Tikochinsky Y, Genislav E, Iragi F, Nave A, Beckmann JS, Gruenbaum Y, Soller M (1990) Large restriction fragments containing poly-TG are highly polymorphic in a variety of vertebrates. Nucleic Acids Res 18:1129–1132

Koorey DJ, Bishop GA, McCaughan GW (1993) Allele non-amplification; a source of confusion in linkage studies employing microsatellite polymorphisms. Hum Mol Genet 2:289–291

Kuhl DP, Caskey CT (1993) Trinucleotide repeats and genome variation. Curr Opin Genet Dev 3:404–407

Kwiatkowski DJ, Henke EP, Weimer K, Ozelius L, Gusella JF, Haines J (1992) Construction of a GT polymorphism map of human 9q. Genomics 12:229–240

Lindahl T (1994) DNA repair. DNA surveillance defect in cancer cells. Curr Biol 4:249–151

Oudet C, Mornet E, Serre JL, Thomas F, Lentes-Zengerling S, Kretz C, Deluchat C, Tejada I, Boue J, Boue A, Mandel JL (1993) Linkage disequilibrium between the fragile X mutation and two closely linked CA repeats suggests that fragile X chromosomes are derived from a small number of founders' chromosomes. Am J Hum Genet 52:297–304

Sanger F, Nicklens S, Carlson AR (1977) DNA sequencing with chain terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schlotterer C, Tautz C (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215

Shibata D, Peinado MA, Ionov Y, Malkhosyan S, Perucho M (1994) Genomic instability in repeated sequences in an early somatic event in colorectal tumorigenesis that persists after transformation. Nat Genet 6:273–281

Sirugo G, Keats B, Fujita R, Duclos F, Purohit K, Koenig M, Mandel JL (1992) Freidreich Ataxia in Louisiana Acadians: demonstration of a founder effect by analysis of microsatellite-generated extended haplotypes. Am J Hum Genet 50:559–566

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–463

Stallings RL (1995) Conservation and evolution of (CT)n/(GT)n microsatellite sequences at orthologous positions in diverse mammalian genomes. Genomics 25:107–113

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilisation of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365:274–276

Todd JA, Aitman TJ, Cornall RJ, Ghosh S, Hall JRS (1991) Genetic analysis of autoimmune type 1 diabetes mellitus in mice. Nature 351:542–547

Urquhart A, Kimpton CP, Gill P (1993) Sequence variability of the tetranucleotide repeat of the beta-actin related pseudogene H-beta-Ac-psi-2 (ACTBP2) locus. Hum Genet 92:637–638

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revised. Genetics 13:737–749

Weber JL, (1990) Informativeness of human (dC-dA)n-(dG-dT)n polymorphisms. Genomics 7:524–530

Weber L, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the Polymerase Chain Reaction. Am J Hum Genet 44:388–396

Weber JL, Wong C (1993) Mutation of human short tandem repeats. Hum Mol Genet 2:1123–1128

Weissenbach J, Gyapay G, Dib C, Vignal A, Morisette J, Millasseau P, Vaysseix G, Lathrop M (1992) A second generation linkage map of the human genome. Nature 359:794–801

Wooster R, Cleton-Jansen AM, Collins N, Mangion J, Cornelis RS, Wiestler OD, Cornelisse CJ, Devilee P, Straton MR (1994) Instability of short tandem repeat (microsatellites) in human cancers. Nat Genet 6:152–156

Zhong N, Dobkin C, Brown WT (1993) A complex mutable polymorphism located within the fragile X gene. Nat Genet 5:248–253.32