

# The tetranucleotide repeat polymorphism D21S1245 demonstrates hypermutability in germline and somatic cells

C. Conover Talbot Jr<sup>1</sup>, Dimitris Avramopoulos<sup>1</sup>, Steven Gerken<sup>2</sup>, Aravinda Chakravarti<sup>3</sup>, John A. Armour<sup>4</sup>, Norisada Matsunami<sup>2</sup>, Ray White<sup>2</sup> and Stylianos E. Antonarakis<sup>1,5,\*</sup>

<sup>1</sup>Center for Medical Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD, <sup>2</sup>Department of Human Genetics, University of Utah, Salt Lake City, UT, <sup>3</sup>Department of Genetics, Case Western Reserve University, Cleveland, OH, USA, <sup>4</sup>Department of Genetics, University of Leicester, UK and <sup>5</sup>The 'Genes 'R' Us' Laboratory of Human Molecular Genetics, Department of Genetics & Microbiology, University of Geneva School of Medicine and Division of Medical Genetics, Cantonal Hospital, Geneva, Switzerland

Received February 16, 1995; Revised and Accepted April 18, 1995

Six novel polymorphic short sequence repeats were identified and localized on the linkage map of human chromosome 21 by genotyping the CEPH reference pedigrees. One of these markers, the tetrameric (AAAG)<sub>n</sub> repeat D21S1245, was found to be hypermutable. In the DNAs from lymphoblastoid cell lines of members of the 40 CEPH families a total of 18 new alleles were detected. These new alleles, sometimes appearing in mosaic forms, arose equally in paternal and maternal DNAs, and could be equally larger or smaller than the alleles from which they were derived. The larger alleles of D21S1245 are more prone to be converted to new alleles. None of the new alleles with mosaicism were present in the corresponding genomic blood DNA, and therefore originated during or after the establishment of the lymphoblastoid cell lines; half of the new alleles without mosaicism were also found in genomic blood DNA of the appropriate CEPH individuals. The range of germline mutation rate observed in the 716 meioses examined was  $0.56\text{--}1.4 \times 10^{-2}$ ; the range of somatic mutations observed in the 405 cell lines examined was  $1.96\text{--}3.46 \times 10^{-2}$ . This is one of the most hypermutable microsatellite repeat polymorphisms in the human genome detected to date. D21S1245, is highly polymorphic (heterozygosity of 0.96) and maps between D21S231 and D21S198.

## INTRODUCTION

The use of DNA polymorphisms has greatly enhanced the ability to map monogenic and complex phenotypes and to explore the human genome in health and disease. In recent years a large number of highly polymorphic loci have been identified in the human genome. These polymorphisms are almost exclusively due to variation of short sequence repeats (SSR) of two to a few nucleotides, whose different alleles can be identified by the polymerase chain reaction (PCR) (1,2).

Extensive linkage maps for all human chromosomes have now been developed and the average resolution of these maps across the entire genome is now less than 5 cM (3–5); for some chromosomes, like chromosome 21, this resolution is less than 2 cM (6,7). The SSRs are thought to evolve very rapidly as many alleles are observed at each locus, leading to their high heterozygosity. Most of the SSRs show new mutant alleles, recognized as expansion or contraction of the number of repeat units at a rate of  $1.2 \times 10^{-3}$  per locus per gamete per generation (8). Such mutation rates are significantly higher for tetranucleotide repeats ( $2.1 \times 10^{-3}$ ) than for dinucleotide repeats ( $0.56 \times 10^{-3}$ ) (8). Notwithstanding, these levels of mutation rates are still low to permit their use as markers in linkage analysis for mapping disease phenotypes. However, a few loci have a much higher mutation rate. For example, Mahtani and Willard (9) reported that the tetranucleotide repeat DXS981 showed a mutation rate of  $1.5 \times 10^{-2}$  (four new mutations in 274 transmissions), which is approximately 10 times higher than the average. As part of our ongoing effort to develop a high-resolution linkage map for chromosome 21 we report here the identification of a highly informative and hypermutable locus, D21S1245. Although this and other hypermutable markers may not be ideal for linkage analyses, they are critical in studies aimed at understanding the instability of DNA markers, such as the genomic instability recognized in the hereditary non-polyposis colon cancer (HNPCC) (10–12), or the specific locus instability as seen by the expansion of trinucleotide repeats associated with some human disorders such as fragile X syndrome (13).

## RESULTS

### Mapping of six SSRs on chromosome 21

A total of six new polymorphic markers were mapped to human chromosome 21; Table 1 shows their map location and some of their characteristics. Two markers were cloned after probing the genomic library with the (GT)<sub>15</sub> oligomer, two with the (AGAT)<sub>5</sub>, and two with the (AAAG)<sub>5</sub> repeat. The heterozygosity of these markers ranges from 31 to 96%. The oligonucleotide primers used for their amplification and

\*To whom correspondence should be addressed at: the University of Geneva Medical School, Centre Medical Universitaire, 1 rue Michel Servet, 1211 Geneva 4, Switzerland

**Table 1.** Characteristics of six new polymorphic markers on human chromosome 21

Marker	Lab name	Repeat type	Oligonucleotides	Genebank #	Size orig STS	Alleles in 8 ceph	Obs het	Close linkage	$\theta$	Lod score
D21S1238	UT409	(CA) <sub>n</sub>	TGCAGAGTGACATGCCTG GTGCTGGATGGATGAAAGG	L16245	217	2	31%	D21S267 D21S259	0.00 0.00	11.14 10.23
D21S1239	UT489	(CA) <sub>n</sub>	ACAATTTCCATGCAGGAAAAC TTGGAAAAACATCAGTCATGAG	L16265	225	5	62%	D21S216 D21S218	0.00 0.00	12.94 10.83
D21S1280	UT706	(AAAG) <sub>n</sub>	AGTGGGCATCAAATTTGGA GAAGAAATAGGAGCCATAAAGC	L31752/3	370	9	94%	D21S224 D21S216	0.02 0.02	25.07 16.15
D21S1244	UT761	(AGAT) <sub>n</sub>	GTCCCCATATTGATAAACTATT ATGAATAGGGGATATGCTGG	L16331	247	5	62%	APP D21S222	0.00 0.00	24.38 22.27
D21S1245	UT762	(AAAG) <sub>n</sub>	GTCAGTATTACCCTGTTACCA GTTGAGGATTTTTGCATCAGT	L16332	289	19	96%	D21S212 D21S198	0.03 0.02	97.94 65.27
D21S1246	UT973	(AGAT) <sub>n</sub>	GATAAAGTAGACAGATAAACA GGATTATAATTCAAGATGAGAT	L16366	400	7	50%	D21S168 D21S231	0.00 0.00	20.47 8.12

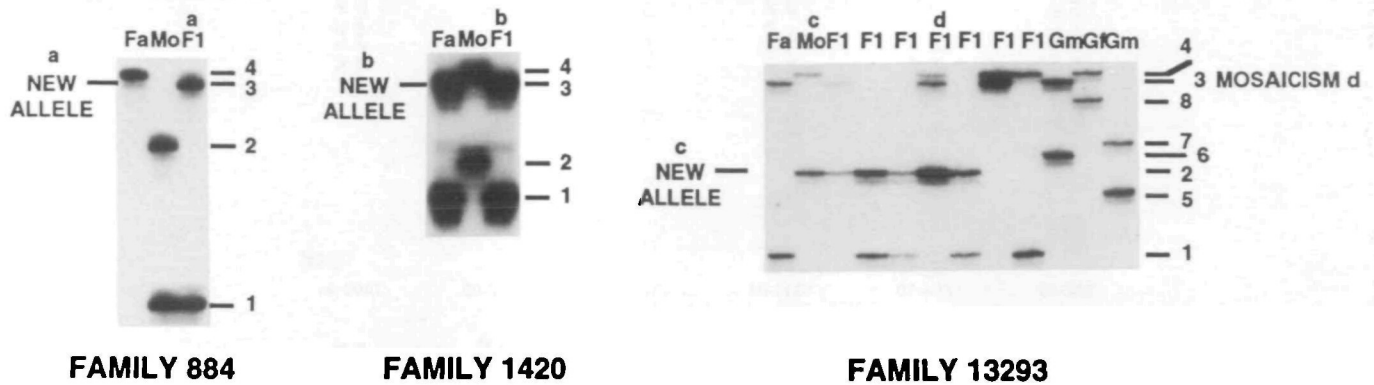


Figure 1. Examples of novel alleles of DNA marker D21S1245. Alleles are numbered per family and the new alleles are shown. For detailed description, see text.

GenBank accession numbers for their nucleotide sequences are included in the table, as are selected two-point lod scores for the most closely linked markers. The locations on the genetic map with odds of 1000:1 against inversion of adjacent loci, relative to markers of the most recent linkage maps of chromosome 21 (6,14) are as follows: D21S1245 maps between D21S231 and D21S198; D21S1246 between D21S156 and D21S168; D21S1238 between D21S1252 and D21S156; D21S1280 and D21S1239 between D21S224 and IFNAR; and D21S1244 between D21S265 and D21S232.

#### Detection of *de novo* alleles for marker D21S1245

One of the polymorphic SSRs, D21S1245, which was identified after screening the library with (AAAG)<sub>5</sub> showed an unusually high number of *de novo* alleles as revealed by initially genotyping the DNAs of lymphoblastoid cell lines of the members of eight CEPH families; the DNAs of all 40 CEPH families were subsequently used for further analysis and mapping. Figure 1 shows examples of *de novo* alleles for marker D21S1245 in DNAs of the CEPH lymphoblastoid cell lines. In family 884, as shown in Fig. 1A, the individual designated F1a has allele 3, which is not present in either her father or mother; the second allele, 1, is inherited from the mother. Non-paternity was excluded in this case with a high degree of certainty as: (i) more than 60 other highly polymorphic markers on chromosome 21 and (ii) hundreds of other SSRs from all other chromosomes have been studied in this family without evidence of non-paternal inheritance of polymorphic alleles. Sample mix-up was also excluded. The most probable explanation for the appearance of allele 3 in this case is the mutation of paternal allele 4 to a shorter allele 3. The size difference between alleles 4 and 3 was four nucleotides, as estimated from acrylamide gel electrophoresis and confirmed by nucleotide sequencing.

In another example (Fig. 1B), the offspring in family 1420, designated F1b, shows alleles 1 and 3, both of which are present in the father; no alleles inherited from the mother were observed. However, uniparental disomy appears to be an unlikely explanation as more than 60 other SSR chromosome 21 markers examined in this family showed biparental inheritance.

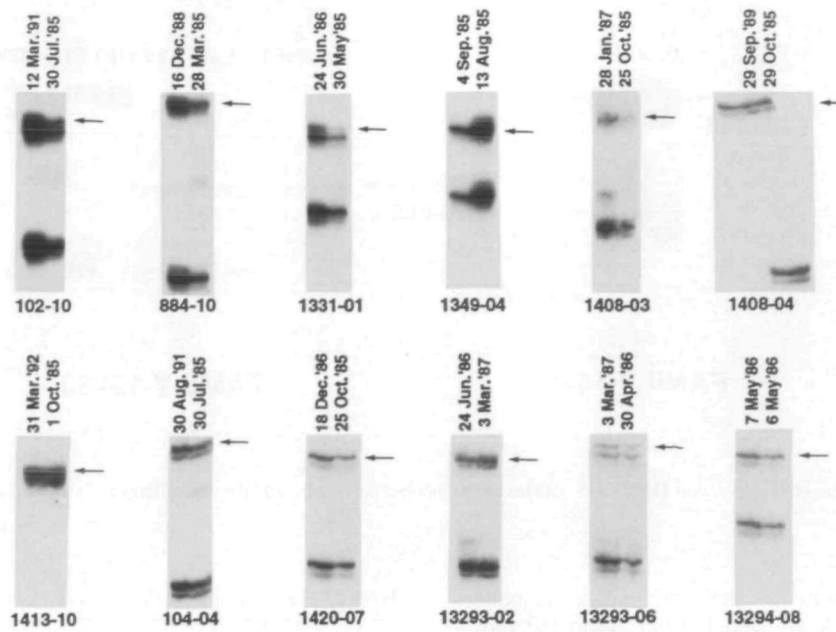
Table 2. Characteristics of *de novo* alleles of marker D21S1245

Paternal source	Paternal	9	50%
	Maternal	9	50%
Size relative to parental	Smaller	8	47%
	Larger	9	53%
Appearance	Single band	13	72%
	Mosaicism	5	28%

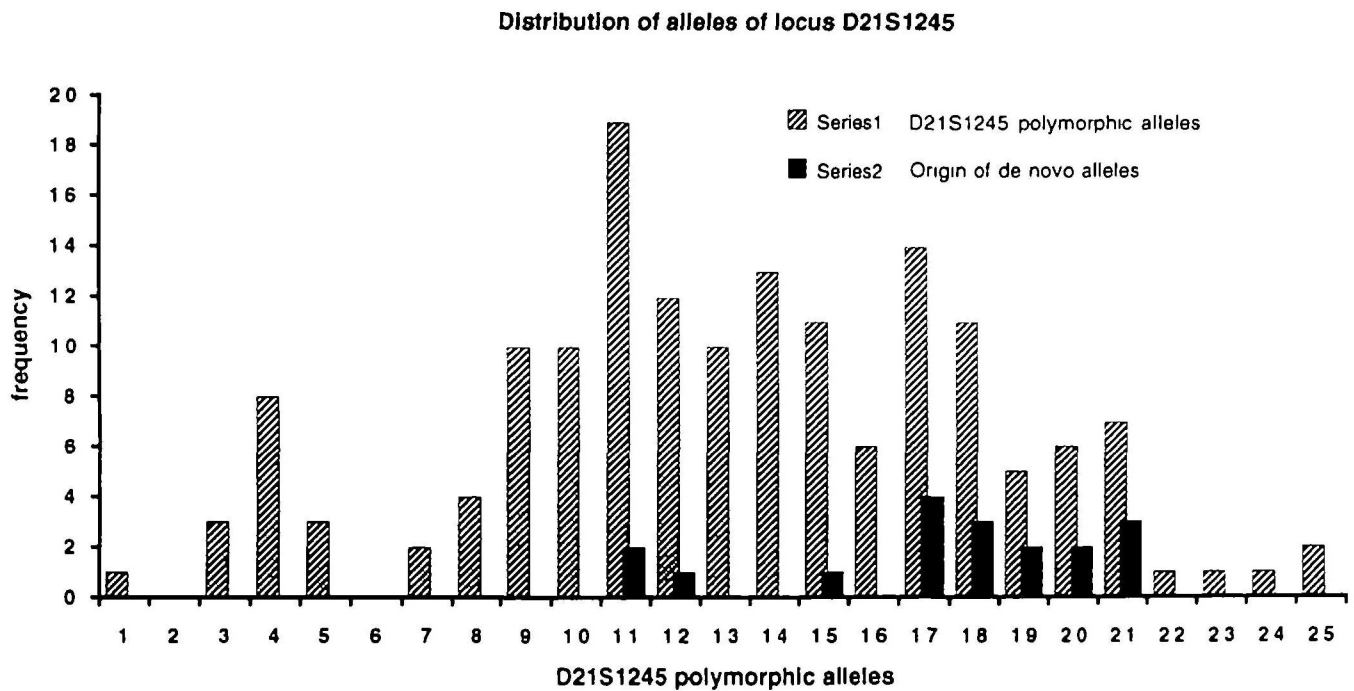
The most likely explanation is that allele 4 of the mother underwent a *de novo* mutation to a new shorter allele 3 which is the same size as paternal allele 3.

In some cases *de novo* alleles occurred as mosaics with their normal counterparts. An example is shown in Fig. 1C. Individual designated F1d from family 13293 shows inheritance of allele 2 from the mother; however, the inherited paternal allele 3 had probably suffered a *de novo* mutation to allele 4 during mitosis so that both of these alleles were present in the DNA of individual F1d. It is noteworthy that in this family allele 2 of the mother was not present in the maternal grandparents, suggesting that allele 2 was probably the result of a mutation in one of the maternal grandmaternal alleles 5 or 7. In this case it is not possible to specify the allele from which the mutation originated.

All members of the 40 CEPH families have been genotyped in the search for *de novo* alleles for marker D21S1245. A total of 18 such *de novo* mutant alleles were observed (DNAs from individuals 13293-02, 13293-06, 13294-08, 1331-01, 1331-07, 1408-03, 1418-01, 1408-04, 1420-07, 1346-05, 1349-07, 1350-06, 1413-10, 104-04, 102-10, 884-10, 37-06, 12-08). Among these, five of 18 (28%) were in mosaic form. Table 2 shows additional characteristics of these events. The origin of the *de novo* mutant alleles was paternal in nine (50%) and maternal in nine (50%) cases, indicating that there was not a preferential sex of origin for these mutations. The size of the *de novo* allele was smaller in eight (47%) and larger in nine (53%) cases. Again, no trend towards expansion or contraction was observed. The parental source of the *de novo*



**Figure 2.** Comparison of PCR results for individuals with new alleles in CEPH DNAs from lymphoblastoid cell lines with different dates of extraction. Samples from the laboratory of S.E.A. are shown on the left and from those kindly supplied from the laboratories of Drs M.Petersen and E.W.Jabs are shown on the right of each pair. The dates of DNA extraction are also shown. Arrows indicate the *de novo* mutant alleles. In the cases of mosaicism (e.g. 13293-06, 1331-01, 102-10), only the novel/mutant allele is shown by arrow.



**Figure 3.** Distribution of alleles of locus D21S1245. The striped bars represent the frequency of each allele in the 80 unrelated parents of the CEPH collection. The black bars represent the frequency distribution of those D21S1245 alleles from which the *de novo* alleles were originated.

allele was assigned by studying the inheritance of alleles in a given pedigree and in some cases by sequence analysis. The parental allele that was closer in size to the *de novo* allele was

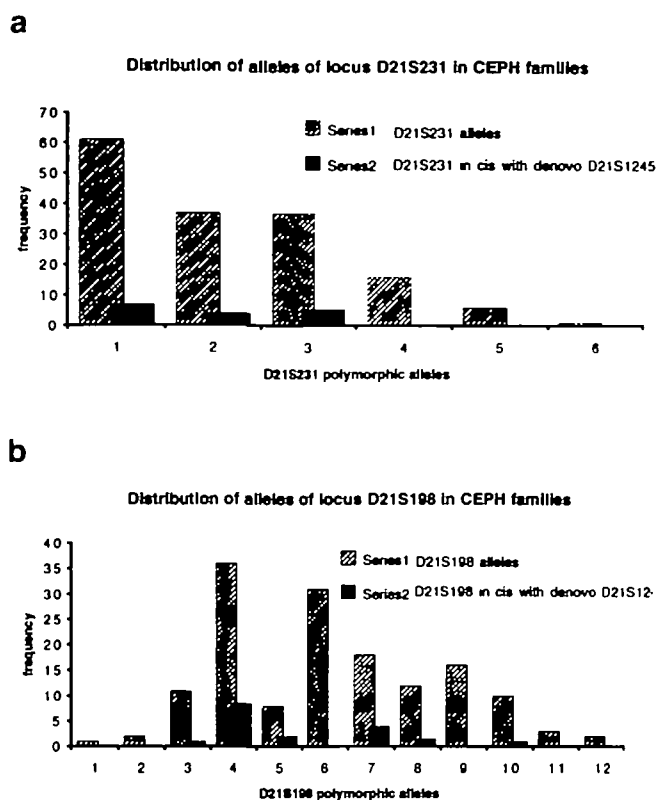
considered to be the one modified to produce the new allele; in a considerable number of cases the assignment was done after sequencing of the individual new allele and its potential

progenitors. Almost all (16 of 18) of the new alleles were checked in the individuals' CEPH DNAs from lymphoblastoid cell lines extracted on two different dates. As shown in Figure 2, no difference was found when DNA samples of the CEPH lymphoblastoid cell lines, extracted on different dates, are compared. In particular, the mosaic cases were similar in both samples (compare for example the similar pattern of lymphoblastoid line DNAs from individual 1331-01 extracted on June 24, 1986 and May 30, 1985).

Mosaicism for *de novo* alleles can result from the establishment or propagation of lymphoblastoid cell lines (8). We therefore compared the D21S1245 genotypes in genomic DNA (samples kindly provided by Dr James Weber, Marschfield, WI) and the corresponding lymphoblastoid cell line DNA in 12 of 18 cases with *de novo* alleles. In the five cases with mosaicism we examined, the *de novo* allele was not found in the genomic DNA, indicating that the *de novo* alleles originated at the establishment or during the propagation of the cell lines. In the seven cases of *de novo* non-mosaicism alleles we examined, the genomic DNA from blood showed the novel alleles in four cases; in the remaining three cases the genomic DNA from blood did not show the *de novo* alleles. Although the data are limited to available blood DNAs, this result suggests that in about half of the non-mosaic cases generation of the *de novo* alleles also occurred in the lymphoblastoid cell lines.

In order to determine whether the likelihood of generation of *de novo* alleles is the same for all sizes of D21S1245 alleles, we compared the distribution of the D21S1245 alleles from all CEPH members with that of those alleles that generated *de novo* forms. This comparison was done because we noticed that the modified allele was almost always the larger one from an individual heterozygote at locus D21S1245. The results of this comparison are shown in Fig. 3. It is apparent that there is a difference between the two distributions of alleles. Most of the *de novo* alleles (>75%) originate from alleles larger than allele 16 (300 nt), while only 30% of the alleles in the CEPH unrelated individuals are larger than allele 16. In fact, classifying alleles by the median *de novo* allele size (#17) shows a highly significant difference ( $\chi^2 = 10.2$ , 1 d.f.,  $0.001 < p < 0.005$ ).

In order to examine the potential contribution of chromosomal sequence environment in the generation of new alleles, we examined the genotypes of D21S231 and D21S198, the two markers on either side and most closely linked to D21S1245, and compared the distribution of their polymorphic alleles with that of those alleles in *cis* with the *de novo* mutations. This was possible because the genetic distances between D21S231-D21S1245 and D21S198-D21S1245 are less than 2 cM; thus polymorphic haplotypes could be developed for these three loci in the CEPH individuals. The histograms for this analysis are shown in Figure 4a,b. Comparison of the histograms by the median *de novo* allele size does not reveal significant differences, suggesting that alleles at the most closely linked loci D21S231 (median = allele 1;  $\chi^2 = 0.04$ , 1 d.f., non-significant) and D21S198 (median = allele 4;  $\chi^2 = 2.67$ , 1 d.f., non-significant) are not involved in the mechanism of the generation of *de novo* alleles. Perhaps the study of additional loci that map closer to D21S1245 will be necessary to rule out involvement of *cis*-acting sequences in the new mutation mechanism(s).



**Figure 4.** (a) Distribution of alleles of locus D21S231 in unrelated individuals of the CEPH families. The striped bars represent the distribution of alleles in the 80 unrelated parents of the CEPH collection. The black bars represent the frequency distribution of those D21S231 alleles in *cis* with the *de novo* D21S1245 alleles. (b) Distribution of alleles of locus D21S198 in unrelated individuals of the CEPH families. The striped bars show the distribution of alleles in the 80 unrelated parents of the CEPH collection. The black bars show the frequency distribution of those D21S198 alleles in *cis* with the *de novo* D21S1245 alleles.

### Sequence analysis of D21S1245 polymorphic alleles

Nucleotide sequence analysis of selected polymorphic alleles from different CEPH DNAs was necessary to determine the origin of the *de novo* allele in some cases and to understand the nature of the polymorphic variability. Sequence analysis revealed that PCR products of the same apparent size are not necessarily the same alleles. The DNA sequence amplified using the D21S1245 oligonucleotide primers contains two units of short sequence repeats; the first unit is a (AAAG)<sub>n</sub> which after an irregular short (GA)<sub>n</sub> area is followed by either another (AAAG)<sub>n</sub> region or by a long (GA)<sub>n</sub> region. Both the tetranucleotide and dinucleotide repeats as well as the connecting region can be polymorphic. It is therefore not surprising that although this SSR was identified by screening with an (AAAG)<sub>n</sub> oligonucleotide probe, the differences among alleles can be two nucleotides rather than four given the polymorphic variability of the (GA)<sub>n</sub> region in certain individuals.

## DISCUSSION

In this study we describe the mapping of six SSR polymorphic loci on human chromosome 21 and the identification of a tetrameric repeat polymorphic marker, D21S1245, that displays a considerable degree of new mutations in the CEPH families. It is important to distinguish between germline and somatic *de novo* mutations in order to derive figures for mutation frequencies. Mutations can be safely classified as germline mutations if they are: (i) clonal (non-mosaic) and (ii) either shown to be present in blood cells, or shown to be in the germline by virtue of transmission (as it is the case in the *de novo* allele in the mother of family 13293 in Fig. 1). Mutations in (i) mosaic form or (ii) not present in blood DNA can be classified as somatic events. As some of the mutations observed in the sample cannot be absolutely categorized, the possible ranges of germline and somatic mutation rates in the 40 CEPH pedigrees are given below. Among the 18 mutations observed there were eight confirmed somatic events and four confirmed germline events. The possible range of germline mutation rate observed in the 716 meioses examined is therefore between four of 716 ( $0.56 \times 10^{-2}$ ) and 10 of 716 ( $1.4 \times 10^{-2}$ ). The possible range of somatic mutations observed in the 405 cell lines examined is between eight of 405 ( $1.96 \times 10^{-2}$ ) and 14 of 405 ( $3.46 \times 10^{-2}$ ).

Weber and Wong (8), studying the mutation rate of SSRs in 28 markers on human chromosome 19, observed an average mutation rate of  $1.2 \times 10^{-3}$  per gamete per generation (approximately 20 000 parent-offspring allele transmissions were studied); tetranucleotide repeats showed a significantly higher average mutation rate ( $2.1 \times 10^{-3}$ ) than did dinucleotide repeats ( $5.6 \times 10^{-4}$ ). Weissenbach *et al.* (3), in their examination of more than  $10^5$  parent-offspring allele transmissions, observed a mutation rate of  $10^{-3}$ . McInnis *et al.* (6) observed a mutation rate of  $0.7 \times 10^{-3}$  for 43 PCR markers on chromosome 21 in their analysis of approximately 15 000 parent-offspring allele transmissions. In the Weber and Wong study (8), the majority of the confirmed new mutations occurred during or after the transformation of lymphoblastoid cells; they observed 21 new mutations *in vivo* and 26 events *in vitro*. In our study, eight of the 18 new mutational events apparently occurred *in vitro* (in somatic cells), four of 18 events *in vivo* (in the germline) and for six the origin of the new mutation could not be determined due to the lack of appropriate blood genomic DNA samples. It is therefore apparent that in the case of locus D21S1245 most of the *de novo* alleles analyzed occurred *in vitro*.

Few SSR polymorphic markers with elevated new mutation rates have been reported. Among those are DXS981, with 1.5% (four of 274 transmissions) (9), and D19S244 with 0.8% new mutation rate (8). Both of these loci were polymorphic due to variable tetranucleotide repeat (GATA)<sub>n</sub> and both had heterozygosities of at least 0.86 with more than 10 alleles each. A short sequence repeat FRAXAC2 that maps within the FMR-1 gene on Xq27.3 has been found to have a new mutation rate of 3.3% but only among fragile X maternal meioses (15); no new mutations in this gene were found in 160 meioses from members of the CEPH pedigrees.

The new mutation rate of two VNTRs (variable number of tandem repeats) exceeds that of the SSRs discussed above. Locus D1S7, comprised of variable numbers of tandem 9

mers, shows a ~5% new mutation rate (16) and locus CEB1 displays a ~15% rate of new mutations, exclusively in male meiosis (17).

In spite of their high degree of heterozygosity, rare polymorphic systems with a high new mutation rate, such as D21S1245 described here, may not be ideal markers for linkage mapping of disease phenotypes, as the appearance of new alleles can complicate interpretation of analyses. On the other hand, such markers may be useful, for example, in evaluating the genomic instability of tumor DNA. Genomic instability for some sporadic and hereditary cancers of the colon, as defined by the appearance of new alleles for short sequence repeat polymorphisms, has been recently recognized (10–12). Subsequently, similar genomic instability has been shown to exist, albeit less commonly, in breast and ovarian tumors as well as in soft tissue sarcomas (18). The use of hypermutable polymorphic marker DXS981, employed in the last study, did not show any advantage over other microsatellite markers.

The generation of new alleles at microsatellite loci is probably due to slipped-strand mispairing (19–21). Although it is currently unknown why some loci are hypermutable, nucleotide sequence comparisons of these loci may reveal sequences that predispose to, or promote the genesis of, new alleles.

The study of the origin of new alleles of marker D21S1245 revealed that the larger alleles were more prone to instability (expansion or contraction). This is similar to the trinucleotide repeat expansion in fragile X syndrome, Huntington disease and similar disorders in which alleles of the premutation size undergo trinucleotide expansion to the full mutation range, whereas alleles of the normal size range do not (13).

## MATERIALS AND METHODS

### Cloning of polymorphic SSR

The cloning of the polymorphic DNA fragments was performed in the laboratory of R.W. as follows: human male DNA was digested with *Sau3A*I, end-filled with dG and dA, and size-fractionated on a 1% agarose gel. DNA fragments ranging from 300 to 700 bp were subcloned into the dC, dT end-filled *Sal*I site of M13mp18. Libraries were screened with end-labeled oligonucleotide probes [e.g. (GT)<sub>15</sub>, (AGAT)<sub>5</sub>, (AAAG)<sub>5</sub>] for recombinant clones containing SSRs. Nucleotide sequences of positive clones were determined using the ABI373 automated sequencer. Oligonucleotide primers were made to develop STSs (sequence tagged sites) for the clones containing long repeats. STS that were polymorphic in eight unrelated CEPH parents were used for genotyping in the CEPH families (22,23). STS were mapped to single human chromosomes using the monochromosomal NIGMS mapping panel 2 (24).

### Linkage mapping

Information on DNA markers that mapped to human chromosome 21 was provided to the laboratory of S.E.A. Genotypes were generated in the DNA of lymphoblastoid cell lines of members of CEPH pedigrees by PCR. The following eight most informative CEPH pedigrees (102, 884, 1331, 1332, 1347, 1362, 1413, 1416) were genotyped for all markers with the exception of locus D21S1245, which was genotyped using the lymphoblastoid cell line DNAs from members of all 40 CEPH pedigrees. PCR amplification was performed as follows: 200 ng of template DNA was amplified in a total volume of 25  $\mu$ l consisting of 1.5 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM Tris (pH 8.3), 200  $\mu$ M of each dNTP, 0.4  $\mu$ M of each primer and 0.5 U *Taq* polymerase. One oligonucleotide primer was <sup>32</sup>P radiolabeled using the T4 kinase reaction. The PCR amplification consisted of an initial denaturation of 94°C for 6 min, followed by 30 cycles of denaturation at 94°C/30 s, annealing at 50–60°C/30 s (varied according to the oligonucleotide pair) and extension at 72°C/30 s using either M.J. Research PTC-100-96 or Perkin-Elmer-Cetus 480 thermocyclers. The PCR products were separated by electrophoresis in 6% poly-

acrylamide/50% urea gel and the dried gels exposed to X-ray film for 2–24 h. Genotypes were determined by two independent readings and entered into our database. In the most recent phase of the project genotypes were determined after electrophoresis of fluorescently labeled PCR products on an ABI373 sequencer. The analysis of the polymorphic alleles was performed using the GENESCAN 672 software (25). Data were transferred electronically to the laboratory of A.C. for linkage mapping. The positions of the new loci in the existing linkage map of chromosome 21 (6) were determined using the program MULTIMAP as described by Matisse *et al.* (5). Locus order and map distances were validated using the program package LINKAGE (26). A more detailed description of the mapping algorithm and the routines for error checking can be found in McInnis *et al.* (6).

Nucleotide sequencing of PCR products of specific alleles of locus D21S1245 was performed as follows: the product of an unlabeled PCR was electrophoresed on a 8% non-denaturing acrylamide gel and the amplified alleles, visualized by ethidium bromide under UV, were separately excised from the gel and recovered by electroelution. A purified single allele's DNA was then used as template for a second round of PCR, the product of which was checked for extra-allelic contamination and purified with a 3%:1% Nusieve/agarose gel and then sequenced using fluorescent dideoxy terminator cycle sequencing on an ABI373 automated DNA sequencer as described (27,28).

## ACKNOWLEDGMENTS

This study was supported by NIH grants HG00468 (to S.E.A.), HD28088 (to A.C.), HG00367 (to R.W.) and Swiss National Foundation grant 31–33965.92 and EU grant OFES 93.0296 to S.E.A. We thank Drs James L. Weber for providing blood genomic DNA from members of the CEPH families and Michael Petersen and Ethylin Jabs for selected lymphoblastoid cell line CEPH DNAs. We also thank A.Scott and the personnel of the Johns Hopkins DNA analysis core facility for their outstanding help in nucleotide sequencing.

## REFERENCES

- Weber, J.L., May, P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.*, **44**, 388–396.
- Litt, M., Luty, J.A. (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.*, **44**, 397–401.
- Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature*, **359**, 794–801.
- Gyapay, G., Morissette, J., Vignal, A., Dib, C., Fizames, C., Millasseau, P., Marc, S., Bernardi, G., Lathrop, M., Weissenbach, J. (1994) *Nature Genet.*, **7**, 246–339.
- Matisse, T.C., Perlin, M., Chakravarti, A. (1994) Automated construction of genetic linkage maps using an expert system (Multimap): a human genome linkage map. *Nature Genet.*, **6**, 384–390.
- McInnis, N.G., Chakravarti, A., Blaschak, J., Petersen, M.B., Sharma, V., Avramopoulos, D., Blouin, J.L., Konig, U., Brahe, C., Cox-Matisse, T., Warren, A.C., Talbot, Jr. C.C., van Broeckhoven, C., Litt, M., Antonarakis, S.E. (1993) A linkage map of human chromosome 32: 43 PCR markers at average intervals of 2.5 cM. *Genomics*, **16**, 562–574.
- Chakravarti, A., Antonarakis, S.E. (1994) Linkage map of human chromosome 21; 75 PCR markers with average interval of 1.5 cM. SCW Chromosome 21, abstract page 1.
- Weber, J.L., Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.
- Mahtani, M.M., Willard, H.F. (1993) A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: implications for mechanisms of mutation at short tandem repeat loci. *Hum. Mol. Genet.*, **2**, 431–437.
- Ionov, J., Peinado, M.A., Malkhosyan, S., Shibata, D., Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
- Thibodeau, S., Bren, G., Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon. *Science*, **260**, 812–816.
- Aaltonen, L.A., Peltonen, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Powell, S.M., Jen, J., Hamilton, S.R., Petersen, G.M., Kinzler, K.W., Vogelstein, B., de la Chapelle, A. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science*, **260**, 812–816.
- Richards, R.I., Sutherland, G.R. (1994) Simple repeat DNA is not replicated simply. *Nature Genet.*, **6**, 114–116.
- Antonarakis, S.E., Chakravarti, A., Blaschak, J., Estivill, X., Avramopoulos, D., Talbot, Jr. C.C., Blouin, J.L., Petersen, M.B., Litt, M., Weissenbach, J., van Broeckhoven, C., Rosen, D., McInnis, M.G., Warren, A.C., Gerken, S. (1993) Linkage map of human chromosome 21; more than 60 PCR markers with average interval of 2.2 cM. HGM93, abstract page 43.
- Zhong, N., Dobkin, C., Brown, W.T. (1993) A complex unstable polymorphism located within the fragile X gene. *Nature Genet.*, **5**, 248–253.
- Jeffreys, A.J., Royle, N.J., Wilson, V., Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature*, **332**, 278–281.
- Vergnaud, G., Mariat, D., Apion, F., Anvias, A., Lathrop, M., Lathier, V. (1991) The use of synthetic tandem repeats to isolate new VNTR loci: cloning of a human hypermutable sequence. *Genomics*, **11**, 135–144.
- Wooster, R., Cleton-Jansen, A.M., Collins, N., Mangion, J., Cornelis, R.S., Cooper, C.S., Gusterson, B.A., Ponder, B.A.J., von Deimling, A., Wiestler O.D., Comelisse, C.J., Devilee, P., Stratton, M.R. (1994) Instability of short tandem repeats (microsatellites) in human cancers. *Nature Genet.*, **6**, 152–156.
- Henderson, S.T., Petes, T.D. (1993) Instability of a plasmid-bore inverted repeat in *S.cerevisiae*. *Genetics*, **14**, 57–62.
- Schlottterer, C., Tautz, D. (1991) Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.*, **20**, 211–215.
- Strand, M., Prolla, T.A., Liskay, R.M., Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting mismatch repair. *Nature*, **365**, 274–276.
- Gerken, S., Whisenant, E., Varkony, T., Todd, S., Gemmill, R., Jones, C., Weissenbach, J., Matsunami, N., Moore, M., Lawrence, E., White, R., Drabkin, H.A. (1994) Physical and genetic mapping of human chromosome 3 loci containing microsatellite repeats. *Chrom. Res.*, **2**, 423–427.
- Murray, J.C., Buetow, K.H., Weber, J.L., Ludwigsen, S., Scherpbier-Heddema, T., Manion, F., Quillen, J., Sheffield, V.C., Sunden, S., Duyk, G.M., Weissenbach, J., Gyapay, G., Dib, C., Morrisette, J., Lathrop, G.M., Vignal, A., White, R., Matsunami, N., Gerken, S., Melis, R., Albertsen, H., Plaetke, R., Odelberg, S., Ward, D., Dausset, J., Cohen, D., Cann, H. (1994) A comprehensive human linkage map with centromorgan density. *Science*, **265**, 2049–2054.
- Drwinga, H.L., Toji, L.H., Kim, C.H., Greene, A.E., Mulivor, R.A. (1992) NIGMS human/rodent somatic cell hybrid mapping panels 1 and 2. *Genomics*, **16**, 311–314.
- Ziegle, J.S., Su, Y., Corcoran, K.P., Nie, L., Mayrand, P.E., Hoff, L.B., McBride, L.J., Kronick, M.N., Diehl, S.R. (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. *Genomics*, **14**, 1026–1031.
- Lathrop G.M., Lalouel J.M. (1984) Easy calculations of lod scores and genetic risks on small computers. *Am. J. Hum. Genet.*, **36**, 460–465.
- Smith, L.M., Sandler, J.Z., Kaiser, R.J., Hughes, P., Dodd, C., Connel, C.R., Heiner, C., Kent, S.B., Hood, L.E. (1986) Fluorescence detection in automated DNA sequence analysis. *Nature*, **321**, 674–679.
- McCombie, W.R., Heiner, C., Kelly, J.M., Fitzgerald, M.G., Gocayne, J.D. (1992) Rapid and reliable fluorescent cycle sequencing of double stranded templates. *DNA Sequence*, **2**, 289–296.

