LETTER

# The Complex Mutation Pattern of a Microsatellite

**Claudia Macaubas,[1,5] Li Jin,[2] Joachim Hallmayer,[2,4] Akinori Kimura,[3] and Emmanuel Mignot[1]**

[1]Center for Narcolepsy, Department of Psychiatry, Stanford University, Palo Alto, California 94304; [2]Department of Genetics, Stanford University, Palo Alto, California 94304; [3]Department of Tissue Physiology, Medical Research Institute, Tokyo Medical and Dental University, Tokyo 101, Japan

*DQCAR* is a $(CA)_n$ microsatellite located in the HLA class II region and tightly linked to *HLA–DQB1*. Previous studies showed a strikingly low level of size variation in *DQCAR* alleles within an extensive subfamily of *HLA–DQ* subtypes (*DQ1*). *DQCAR* alleles in non-*DQ1* subtypes showed a higher degree of size polymorphism. In this study sequence analysis demonstrates that *DQ1*- associated *DQCAR* alleles have a single C → A nucleotide substitution interupting the CA repeat array. Frequent CA → GA mutations are also observed in *DQ1*-associated microsatellites with identical allele sizes. In contrast, *DQCAR* alleles associated with non-*DQ1* haplotypes display a perfect CA repeat sequence and the variation in allele size is attributable only to differences in the number of CA repeats. Our results imply that several mutational mechanisms are involved in the generation of allelic diversity within the same microsatellite locus. The possibility of different mutation rates in the same locus should to be taken into account when using these markers in evolutionary and disease studies.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. U96944–U96962 and S87165.]

Microsatellites are abundant in mammalian genomes (Weber 1990; Stallings et al. 1991) and have attracted an enormous amount of interest in recent years. Instability of trinucleotide repeats is involved with certain neurodegenerative conditions (for review, see Rosenberg 1996), suggesting that underlying mutational mechanisms may be responsible for genetic morbidity. Polymorphic dinucleotide repeats have been used extensively for gene mapping (Dib et al. 1996), forensic identification (Edwards et al. 1992) and studying the history of human population (Bowcock et al. 1994; Deka et al. 1995).

In spite of their popularity as a genetic tool, the mechanisms underlying microsatellite allelic diversity are still poorly understood. Estimations of mutation rate for microsatellites range between $10^{-2}$ and $10^{-5}$ per generation (Edwards et al. 1992; Mahtani and Willard 1993), with Weber and Wong (1993) observing that dinucleotide repeats have a lower average mutation rate than tetranucleotides ($5.6 \times 10^{-4}$ and $2.1 \times 10^{-3}$, respectively). Most of the observed changes in microsatellite allele sizes are attributable to deletion or addition of a few repetitive units (Straub et al. 1993; Deka et al. 1995). In principle, this pattern of mutation fits with a simple stepwise mutation model (Ohta and Kimura 1973) but might not account for all the variation observed (Shriver et al. 1993; Valdes et al. 1993; Deka et al. 1995). A two-phase model that allows for bigger differences in repetitive units has also been proposed (Di Rienzo et al. 1994). Strand slippage during replication has been suggested as the most likely mechanism involved in the generation of microsatellite mutations (Levinson and Gutman 1987; Schlötterer and Tautz 1992; Strand et al. 1993).

Most of the studies involving microsatellites have analyzed a large number of loci. Detailed sequence analyses of mutation patterns at individual loci have been performed especially for trinucleotides associated with diseases (Rosenberg 1996) and some tetranucleotides (Mahtani and Willard 1993; Talbot et al. 1995). Minisatellites have also been analyzed (Jeffreys et al. 1994). Dinucleotides, however, have received less attention.

Previously, we studied a $(CA)_n$ microsatellite (*DQCAR*) located in the human leukocyte antigen (HLA) class II region, ~12 kb centromeric to *HLA–DQA1* and 1–2 kb telomeric to the *HLA–DQB1* gene

[4]Present address: Graylands/UWA Centre for Clinical Research in Neuropsychiatry, Gascoyne House, Graylands Hospital, WA 6010, Australia.
[5]Corresponding author. Present address: TVW Telethon Institute for Child Health Research, West Perth, WA 6872, Australia.
E-MAIL macaubas@ichr.uwa.edu.au; FAX 61-9-388-3414.

(Satyanarayana and Strominger 1992). *DQA1* and *DQB1* encode for the respective α and β chains, which form the heterodimeric protein *HLA–DQ* expressed in the surface of cells involved in the immune response. Fourteen different *DQCAR* alleles were identified and these were found to be in tight linkage disequilibrium with *HLA–DQ* (Macaubas et al. 1995; Mignot et al. 1995; Jin et al. 1996). Extremely low size polymorphism was found within *DQCAR* alleles associated with *DQ1*, the predominant *HLA–DQ* subtype in several populations (Imanishi et al. 1992). The *DQCAR* repeat associated with the other DQ subtypes (*DQ2*, *DQ3*, and *DQ4*) showed both larger numbers and sizes of alleles (Macaubas et al. 1995). In this study we carried out a sequence analysis of *DQCAR* alleles to investigate the basis for the observed allelic differences between *DQ1* and non-*DQ1* haplotypes.

## RESULTS

The phylogeny of *DQB1* alleles is shown in Figure 1 (left). There are two groups of alleles—the *DQ1* subgroup (*DQB1*05* and *DQB1*06*) and non-*DQ1* (*DQB1*02*, *DQB1*03*, and *DBQ1*04*). The typing of 2140 chromosomes (Fig. 1, column 2), showed that in the non-*DQ1* subfamily, more *DQCAR* alleles per DQB1 subtypes were found (Fig. 1, column 3) in comparison with the *DQ1*-associated subgroup. Non-*DQ1*-associated *DQCAR* alleles also have longer microsatellite sizes (Macaubas et al. 1995; Jin et al. 1996). The exceptions were *DQCAR* alleles associated with *DQB1*0201*, with only two short microsatellite alleles detected in >150 chromosomes typed.

At least one *DQCAR* allele from 15 *DQB1* lineages was then subcloned and sequenced. All *DQCAR* alleles associated with the *DQ1* subfamily showed a C → A nucleotide substitution in the fifteenth base pair, resulting in a CAAA motif interrupting the CA repeat array (Table 1; Fig. 2). In contrast, the microsatellite allele sequences associated with non-*DQ1* subfamilies showed an uninterrupted CA repeat motif (Table 1; Fig. 2).

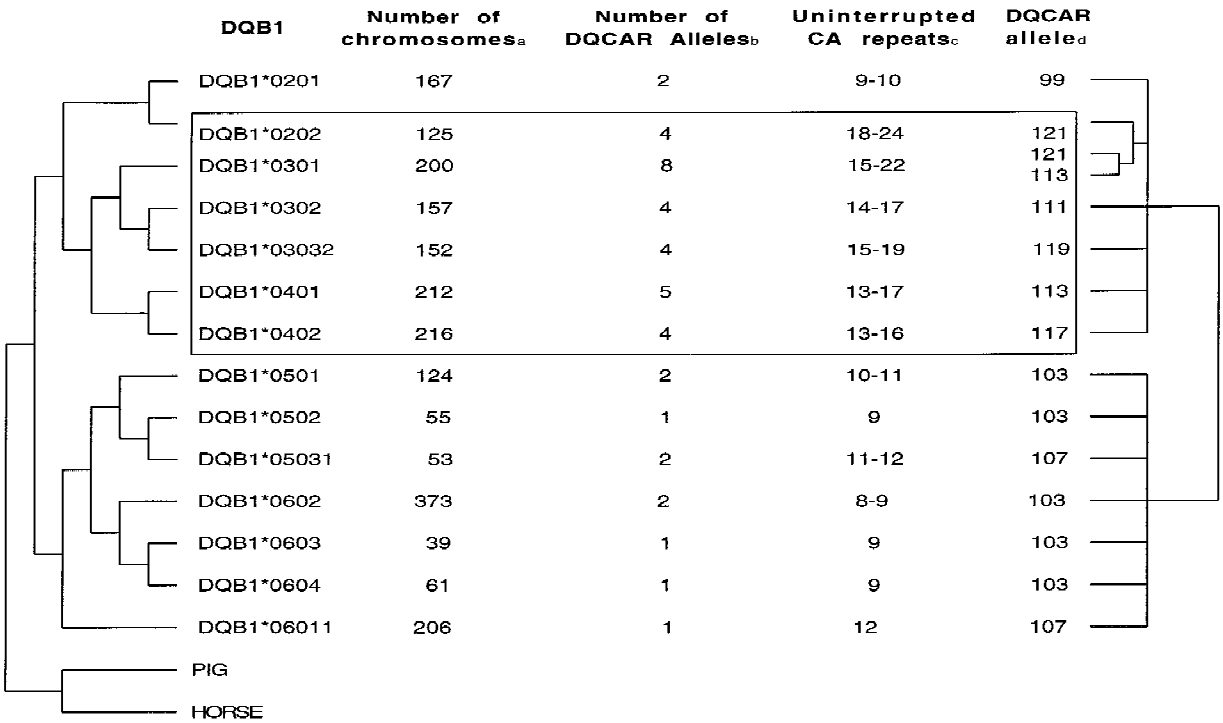


Figure 1 Phylogenetic relationships of *DQB1* alleles (*left*) and *DQCAR* alleles (*right*) based on their sequence data. Phylogenies of 14 *DQB1* and 15 *DQCAR* alleles were reconstructed as detailed in Methods. (*a*) Number of chromosomes typed for *DQB1* and *DQCAR* in a mixed population. (*b*) Number of *DQCAR* alleles associated with each *DQB1* allele in the mixed population. (*c*) Number of uninterrupted CA repeats for each *DQCAR* allele observed in the mixed population, deduced from the sequenced alleles on Table 1. (*d*) The *DQCAR* alleles represented in the tree are the sequenced from Table 1. Areas with high *DQCAR* size variation (more than two *DQCAR* alleles found in association with a lineage) are boxed.

Table 1. ***DQCAR*** Sequences Observed in Association with Various ***HLA–DQB1*** Alleles

| DQ | DQB1 | Cell line | DQCAR allele[a] | Sequence | Repeat motif |
|---|---|---|---|---|---|
| | | | | 10 20 30 40 50 60 | |
| 1(6) | *0602 | 9010 | 103 | CAGACACACA CACAACACA CACACACACA CA-------- ---------- GAGAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)8(GA)4 |
| 1(6) | *0602 | 9083 | 103 | CAGACACACA CACAACACA CACACACACA CA-------- ---------- GAGAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)8(GA)4 |
| 1(6) | *0603 | 9062 | 103 | CAGACACACA CACAACACA CACACACACA CACA------ ---------- --GAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)9(GA)3 |
| 1(6) | *0604 | 9063 | 103 | CAGACACACA CACAACACA CACACACACA CACA------ ---------- --GAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)9(GA)3 |
| 1(6) | *0609 | 9055 | 103 | CAGACACACA CACAACACA CACACACACA CACA------ ---------- --GAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)9(GA)3 |
| 1(5) | *0502 | 9012 | 103 | CAGACACACA CACAACACA CACACACACA CACA------ ---------- --GAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)9(GA)3 |
| 1(5) | *0501 | 9080 | 103 | CAGACACACA CACAACACA CACACACACA CACACA---- ---------- ----GAGATG AGGTAAGA | CAGA(CA)5AA(CA)10(GA)2 |
| 1(5) | *0501 | 9004 | 103 | CAGACACACA CACAACACA CACACACACA CACACA---- ---------- ----GAGATG AGGTAAGA | CAGA(CA)5AA(CA)10(GA)2 |
| 1(5) | *05031 | 9054 | 107 | CAGACACACA CACAACACA CACACACACA CACACACA-- ---------- --GAGAGATG AGGTAAGA | CAGA(CA)5AA(CA)11(GA)3 |
| 1(6) | *0601 | 9011 | 107 | CAGACACACA CACAACACA CACACACACA CACACACACA ---------- ----GAGATG AGGTAAGA | CAGA(CA)5AA(CA)12(GA)2 |
| 2 | *0201 | 9019 | 99 | CAGACACACA CACACACACA CA-------- ---------- ----GCAAGA GAGAGAGATG AGATAAGA | CAGA(CA)9 GCAA(GA)5 |
| 3(8) | *0302 | 9029 | 111 | CAGACACACA CACACACACA CACACACACA CACA------ ----GCAAGA GAGAGAGATG AGATAAGA | CAGA(CA)15 GCAA(GA)5 |
| 3(8) | *0302 | 9034 | 111 | CAGACACACA CACACACACA CACACACACA CACA------ ----GCAAGA GAGAGAGATG AGATAAGA | CAGA(CA)15 GCAA(GA)5 |
| 3(7) | *0301 | 9070 | 113 | CAGACACACA CACACACACA CACACACACA CACAGA---- ----GCAAGA GAGAGAGATG AGATAATA | CAGA(CA)15GA GCAA(GA)5 |
| 4 | *0401 | N°106 | 113 | CAGACACACA CACACACACA CACACACACA CACACA---- ----GCAAGA GAGAGAGATG AGATAAGA | CAGA(CA)16 GCAA(GA)5 |
| 4 | *0402 | 9072 | 115[1] | CAGACAGACA CACACACACA CACACACACA CACACACA-- ----GCAAGA GAGAGAGATG AGATAAGA | (CAGA)2(CA)15 GCAA(GA)5 |
| 3(9) | *03032 | 9052 | 117[1] | CAGACACACA CACACACACA CACACACACA CACACACACA ----GCAAGA GAGAGAGATG AGATAAGA | CAGA(CA)18 GCAA(GA)5 |
| 3(7) | *0301 | 9035 | 119[1] | CAGACACACA CACACACACA CACACACACA CACACACACA CA--GCAAGA GAGAGAGATG AGATAATA | CAGA(CA)19 GCAA(GA)5 |
| 2 | *02 | MANN | 121 | CACACACACA CACACACACA CACACACACA CACACACACA CACAGCAAGA GAGAGAGATG AGATAATA | (CA)22 GCAA(GA)5 |
| 2 | *0202 | 9051 | 121 | CACACACACA CACACACACA CACACACACA CACACACACA CACAGCAAGA GAGAGAGATG AGATAATA | (CA)22 GCAA(GA)5 |

Sequences were obtained as described in Methods.
[a](1) Actual size of the fragment sequenced (2 bp less than the allele defined initially; Macaubas et al. 1995). *DQCAR* primers sequence are also included in the size of the allele.

Non-*DQ1*-associated alleles differed in size as a result of varying numbers of CA repeats. *DQCAR 99* (the shortest allele sequenced) has a repeat array of 9 uninterrupted CA units, and *DQCAR 121* (the longest) has a stretch of 22 CA units. Some changes from C → G into the repeat array were also observed in some alleles on both sides of the CA repeat (Table 1).

Each *DQCAR* allele is composed of a CA repeat, followed by a GA motif. Non-*DQ1*-associated alleles always show an intervening tetramer GCAA between CA and GA motifs. In *DQ1*-associated alelles the GA repeat structure lies immediately next to the CA repeat array (Table 1; Fig. 2). The GA repeat motif in non-*DQ1*-associated alleles is monomorphic, with five repeats of GA (Table 1; Fig. 2). The number of GA repeats varies for *DQ1*-associated alleles, even when the sizes of two alleles, are identical. For example, *DQCAR* 103 alleles associated with *DQB1*0602* and with *DQB1*0603* have the motif (CA)8(GA)4 and (CA)9(GA)3, respectively. The total number of repetitive units for those alleles remains constant (i.e., 12 units) while their composition varies. Using routine genotyping methods, this variation is missed.

The sequence flanking the 3′ end of *DQCAR* showed subtype-specific variations. *DQ1*-associated alleles showed a G nucleotide in the fifth nucleotide after the GA run (Table 1; Fig. 2). Non-*DQ1* alleles have an A in this position. Sequences from *DQB1*0202* and *DQB1*0301* reveal a T in the ninth nucleotide after the GA stretch. All of the other sequences (including *DQ1* sequences) have a G in this position (Table 1).

A phylogenetic tree of *DQCAR* alleles, based on their flanking sequence, was then constructed (Fig. 1, right). As for *DQB1* alleles, two groups were observed: *DQCAR* alleles associated with *DQ1* subtypes formed one group, whereas microsatellite alleles observed with non-*DQ1* subtypes clustered in a separate group. The number of uninterrupted CA repeats for each *DQCAR* allele observed in the typed population was deduced from the sequence data (Fig. 1, column 4). The non-*DQ1*-associated *DQCAR* alleles showed a higher number of uninterrupted CA repeats (13–24 CA repeats), compared to *DQCAR* associated with *DQ1* haplotypes (between 8 and 12 CA repeats). The only exceptions were *DQCAR* alleles associated with *DQB1*0201*, with 9 and 10 repeats.

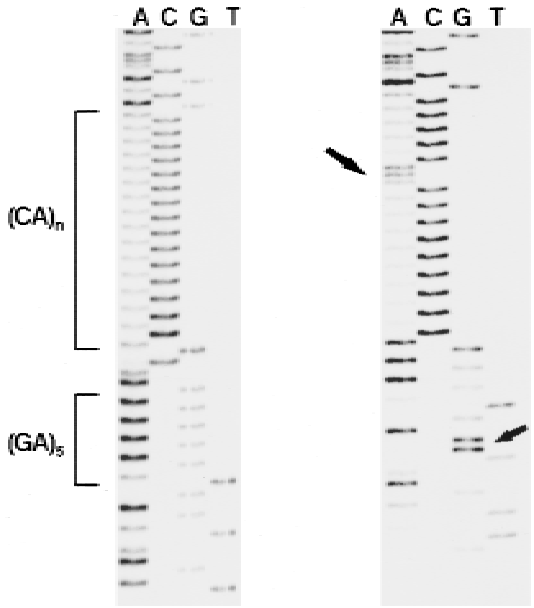


Figure 2 Sequence from two representative cell lines. (Left) Cell line 9072 (DQ4, *DQB1*0402, DQCAR117*); (right) cell line 9062 (*DQ1, DQB1*0603, DQCAR103*). The left arrow points to the A nucleotide interrupting the CA repeat sequence in the *DQ1* cell line. Note the tetramer GCAA between the $(CA)_n$ region and the (GA)5 in the DQ4 cell line. The right arrow points to the G substitution in the flanking region, characteristic of *DQCAR103* and *DQCAR107*. Sequence of opposite strand confirmed the results in both cases (not shown). Sequence was performed as detailed in Methods.

## DISCUSSION

In the *DQCAR* locus, three forms of mutations are responsible for the generation of new alleles: (1) simple gain or loss of CA units; (2) a single C → A substitution in the CA repeat array; and (3) GA → CA or CA → GA mutations in the region immediately following the CA repeat array. Only the first mutation mechanism resulted in a change of allele size.

We observed that for *DQCAR* alleles with perfect CA-repeat runs, the change in allele size is attributable to small differences in the number of repetitive units, mostly by only 2 bp. This pattern of mutation seems to fit with a model of small changes in repeat unit number via a strand-slippage mechanism (Strand et al. 1993). In addition, the increase was more remarkable above 13 repeat units, suggesting, as pointed out by Weber (1990), that informativeness tends to increase with the number of repeats.

The interruption of the perfect repeat array had a dramatic effect on the stability of the *DQCAR* alleles. It has been shown that imperfect repeats are less polymorphic (Weber 1990; Garza et al. 1995), and, interestingly, stabilization of repetitive sequences has also been described for certain trinucleotide repeats associated with diseases. In spinocerebellar ataxia type I (SCA1), Chung et al. (1993) showed that expanded alleles had a continuous run of $(CAG)_n$ repeats in the SCA1 region, whereas unexpanded alleles had CAT interruptions. The gene for spinocerebellar ataxia 2 (SCA2) has been identified recently and also contains a stretch of $(CAG)_n$ repeats uninterrupted in mutated alleles and interrupted by one to three CAAs in normal ones (Imbert et al. 1996). Similarly, fragile X syndrome is associated with an expansion of an almost uninterrupted run of triple repeat $(CGG)_n$ at the 5′-untranslated region of the *FMR1* gene with stable alleles showing interspersed AGGs (Eichler et al. 1994; Zhong et al. 1995).

It is not totally understood how interruptions stabilize repetitive sequences. Interruptions may create more matching points during replication, and a misalignment could be detected more easily by mismatch repair enzymes, as suggested by Heale and Petes (1995). They observed that in yeast, the interruption of a repetitive run by one different repeat increased by 100 times its stability, and this stabilization was dependent on a functional DNA mismatch repair system.

Interestingly, the phylogeny of *DQB1/DQCAR* haplotypes suggests that the C → A mutation interrupting the CA repeat is monophyletic and therefore occurred only once in the evolution of this locus. The interruption of the CA repeat array in *DQ1*-associated *DQCAR* alleles must have been present before the divergence of this subtype, as it was found in all *DQCAR* alleles associated with *DQ1* haplotypes and in none of the non-*DQ1* haplotypes. The four major DQ subfamilies (*DQ1, DQ2, DQ3,* and *DQ4*) can be identified in nonhuman primates. They have persisted for at least 5 million years, and diversity within the subfamilies seems to have arisen after speciation (Gyllensten et al. 1990).

Despite the low level of size variations in the *DQCAR* alleles associated with *DQ1* haplotypes, additional sequence diversity was observed between alleles of identical size because of CA → GA changes at the end of the repeat. As the number of CA units increases, a concomitant decrease in the number of GA units keeps the overall allele size constant. Additional polymorphism is thus obtained for alleles of similar size, as observed previously at other microsatellite loci (Saha et al. 1993; Estoup et al. 1995).

Sequence variation has also been observed in minisatellites (Jeffreys et al. 1994). The variation in the repetitive sequence is not restricted to *DQCAR* alleles associated with *DQ1* alleles but is also found in alleles associated with non-*DQ1* haplotypes (e.g., *DQB1*0301* and *DQB1*0402*). Nevertheless, the frequent change of allelic composition of CA and GA repeat numbers in alleles of identical size suggests a new mutational mechanism that has not been reported previously in microsatellites.

In conclusion, sequence analysis shows a complex mutational pattern in a single $(CA)_n$ microsatellite. This complexity could also occur in other microsatellite loci and must be considered when using these markers for evolutionary and gene mapping studies.

## METHODS

### Tree Construction

A phylogeny of 14 frequent *DQB1* alleles was reconstructed on the basis of the aligned coding sequences of these alleles using the neighbor-joining method (Saitou and Nei 1987). Genetic distances were estimated using Kimura's two-parameter model (Kimura 1980). Several other measures of genetic distances were used, and all gave identical results in terms of the topology of the phylogeny obtained using the neighbor-joining method (Jin et al. 1996). The maximum parsimony tree displayed a slightly different topology, which became identical with the neighbor-joining tree once peptide-binding sites were removed from the analysis. The tree was rooted by DQB sequences of pig and horse. A maximum parsimony tree of *DQCAR* locus based on the flanking sequences of *DQCAR* alleles was also constructed.

### Microsatellite Amplification and Sequence

Genomic DNA (200 ng) was used as a template, in a total volume of 25 µl of 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 50 pmoles of each primer (see below), 125 µM of each dNTP, and 1 unit of AmpliTaq DNA polymerase (Perkin Elmer). After an initial incubation for 5 min at 95°C, the amplification was carried for 30 cycles, with each cycle consisting of 30 sec at 94°C, 1 min at 57°C, and 1 min at 72°C, with a final extension at 72°C for 15 min. The primers used were CAR1, 5′-GAAACATATATTAACAGAGACAGACAAA-3′; and CAR2, 5′-CATTTCTCTTCCTTATCACTTCATA-3′ (Satyanarayana and Strominger 1992; Macaubas et al. 1995). The presence of PCR product was confirmed by gel electrophoresis, followed by extraction with chloroform and ligation to pT7Blue T-vector (Novagen), as detailed in the product manual. Both strands of plasmid DNA from several clones were sequenced with Sequenase version 2.0 T7 DNA Polymerase (U.S. Biochemical), using T7 and U19 primers, following the manufacturer's instructions

### Cell Lines

Twenty HLA class II homozygous B cell lines were analyzed. Eighteen B cell lines were obtained from the Tenth Histocompatibility Workshop panel (Yang et al. 1987; Kimura et al. 1992; Aldener and Olerup 1993). The other two B cell lines were cell line MANN (donated by Dr. George Blanck, Harvard University, Cambridge, MA; Blanck and Strominger 1988) and B cell line no. 106, acquired from the UCLA DNA reference panel (Dr. P.I. Terasaki, University of California, Los Angeles).

### HLA Typing

Chromosomes (2140) from several populations were typed for *HLA–DQB1*, *HLA–DQA1* and *DQCAR* loci, and haplotypes were deduced as described in Kimura et al. (1992) and Jin et al. (1996).

## ACKNOWLEDGMENTS

This work was funded by National Institutes of Health (NIH) grants NS23724 and 33797 (to E.M.) and by NIH training grant T32-GS08404 (to L.J.) We thank Drs. P. Underhill and P. Oefner for suggestions regarding cloning and sequencing strategies, and Professor Jonathan Majer for reviewing the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked ''advertisement'' in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

Aldener, A. and O. Olerup. 1993. Characterization of a novel DQB1 (DQB1*0609) allele by PCR amplification with sequence-specific primers (PCR-SSP) and nucleotide sequencing. *Tissue Antigens* 42: 536–538.

Blanck, G. and J.L. Strominger. 1988. Molecular organization of the DQ subregion (DO-DX-DV-DQ) of the human MHC and its evolutionary implications. *J. Immunol.* 141: 1734–1737.

Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J.R. Kidd, and L.L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368: 455–457.

Chung, M., L.P.W. Ranum, L.A. Duvick, A. Servadio, H.Y. Zoghbi, and H.T. Orr. 1993. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellat ataxia type I. *Nature Genet.* 5: 254–258.

Deka R., L. Jin, M.D. Shriver, L.M. Yu, S. DeCroo, J. Hundrieser, C.H. Bunker, R.E. Ferrel, and R. Chakraborty. 1995. Population genetics of dinucleotide $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms in world populations. *Am. J. Hum. Genet.* 56: 461–474.

Dib, C., S. Faure, C. Fizames, D. Samson, N. Drouot, A. Vignal, P. Millasseau, S. Marc, J. Hazan, E. Seboun, M. Lathrop, G. Gyapay, J. Morissette, and J. Weissenbach. 1996. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 380: 152–154.

DiRienzo, A., A.C. Peterson, J.C. Garza, A.M. Valdes, M. Slatkin, and N.B. Freimer. 1994. Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci.* 91: 3166–3170.

Edwards, A., H.A. Hammond, L. Jin, C.T. Caskey, and R. Chakraborty. 1992. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Proc. Natl. Acad. Sci.* 91: 3166–3170.

Eichler, E.E., J.J.A. Holden, B.W. Popovich, A.L. Reiss, K. Snow, S.T. Thibodeau, C.S. Richards, P.A. Ward, and D. L. Nelson. 1994. Length of uninterrupted CGG repeats determines instability in the *FMR1* gene. *Nature Genet.* 8: 88–94.

Estoup, A., C. Tailliez, J.-M. Cornuet, and M. Solignac. 1995. Size homoplasy and mutational processes of interrupted microsatellites in two bee species, *Apis mellifera* and *Bombus terrestris* (Apidae). *Mol. Biol. Evol.* 112: 1074–1084.

Garza, J.C., M. Slatkin, and N.B. Freimer. 1995. Microsatellite allele frequencies in human and chimpazees, with implications for constraints on allele size. *Mol. Biol. Evol.* 12: 594–603.

Gyllensten, U.B., D. Lashkari, and H.A. Erlich. 1990. Allelic diversification at the class II DQB locus of the mammaliam major histocompatibility complex. *Proc. Natl. Acad. Sci.* 87: 1835–1839.

Heale, S.M. and T.D. Petes. 1995. The stabilization of repetitive tracts of DNA by variant repeats requires a functional DNA mismatch repair system. *Cell* 83: 539–545.

Imanishi, T., T. Akaza, A. Kimura, K. Tokunaga, and T. Gojobori. 1992. Allele and Haplotype frequencies for HLA and complement loci in various ethnic groups. In *Proceedings of the Eleventh International Histocompatibility Workshop and Conference* (ed. K. Tsuji, M. Aizawa, and T. Sasazuki), pp. 1065–1220. Oxford Science Publications, New York, NY.

Imbert, G., F. Saudou, G. Yvert, D. Devys, Y. Trottier, J.-M. Garnier, C. Weber, J.L. Mandel, G. Cancer, N. Abbas, A. Dürr, O. Didierjean, G. Stevanin, Y. Agid, and A. Brice. 1996. Cloning for the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genet.* 14: 285–291.

Jeffreys, A.J., K. Tamaki, A. MacLeod, D.G. Monckton, D.L. Neil, and J.A.L. Armour. 1994. Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.* 6: 136–145.

Jin, L., C. Macaubas, J. Hallmayer, A. Kimura, and E. Mignot. 1996. Mutation rate varies among alleles at a microsatellite locus: Philogenetic evidence. *Proc. Natl. Acad. Sci.* 93: 15285–15288.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16: 111–120.

Kimura, A., R.-P. Dong, H. Harada, and T. Sasazuki. 1992. DNA typing of HLA class II genes in B-lymphoblastoid cell lines homozygous for HLA. *Tissue Antigens* 40: 5–12.

Levinson, G. and G.A. Gutman. 1987. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203–221.

Macaubas, C., J. Hallmayer, J. Kalil, A. Kimura, S. Yasunaga, F.C. Grumet, and E. Mignot. 1995. Extensive polymorphism of a (CA)n microsatellite located in the HLA-DQA1/DQB1 class II region. *Hum. Immunol.* 42: 209–220.

Mahtani, M.M. and H.F. Willard. 1993. A polymorphic X-linked tetranucleotide repeat locus displaying a high rate of new mutation: Implications for mechanisms of mutation at short tandem repeat. *Hum. Mol. Genet.* 2: 431–437.

Mignot, E., A. Kimura, M. Abbal, E. Thorsby, X. Lin, A. Voros, C. Macaubas, F. Bouissou, L.M. Sollid, A. Cambon-Thomsen, and F.C. Grumet. 1995. DQCAR microsatellite polymorphisms in three selected HLA class II-associated diseases. *Tissue Antigens* 46: 299–304.

Ohta, T. and M. Kimura. 1973. The model of mutation appropiate to estimate the number of electrophoretically detectable alleles in a genetic population. *Genet. Res.* 22: 201–204.

Rosenberg, R.N. 1996. DNA-triplet repeats and neurologic disease. *N. Engl. J. Med.* 335: 1222–1224.

Saha, B.K., J.J. Shields, R.D. Miller, T.H. Hansen, and D.C. Shreffler. 1993. A highly polymorphic microsatellite in the class II *Eb* gene allows tracing of major histocompatibility complex evolution in mouse. *Proc. Natl. Acad. Sci.* 90: 5312–5316.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4: 406–425.

Satyanarayana, K. and J.L. Strominger. 1992. DNA sequences near a meiotic recombinational breakpoint within the human *HLA-DQ* region. *Immunogenetics* 35: 235–240.

Schlötterer, C. and D. Tautz. 1992. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* 20: 211–215.

Shriver, M.D., L. Jin, R. Chakraborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics* 134: 983–993.

Stallings, R.L., A.F. Ford, D. Nelson, D.C. Torney, C.E. Hildebrand, and R.K. Moyzis. 1991. Evolution and distribution of $(GT)_n$ repetitive sequences in mammaliam genomes. *Genomics* 10: 807–815.

Strand, M., T.A. Prolla, R.M. Liskay, and T.D. Petes. 1993. Desestabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365: 274–276.

Straub, R.E., M.C. Speer, Y. Luo, K. Rojas, J. Overhauser, J. Ott, and T.C. Gilliam. 1993. A microsatellite genetic linkage map of human chromosome 18. *Genomics* 15: 48–56.

Talbot, C.C., Jr., D. Avramopoulos, S. Gerken, A. Chakravarti, J.A. Armour, N. Matsunami, R. White, and S.E. Antonarakis. 1995. The tetranucleotide repeat D21S1245 demonstrates hypermutability in germline and somatic cells. *Hum. Mol. Genet.* 4: 1193–1199.

Valdes, A.M., M. Slatkin, and N.B. Freimer. 1993. Allele frequencies at microsatellite loci: The stepwise mutation model revisited. *Genetics* 133: 737–749.

Weber, J.L. 1990. Informativeness of human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ polymorphisms. *Genomics* 7: 524–530.

Weber, J.L. and C. Wong. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* 2: 1123–1128.

Yang, S.Y., E. Milford, U. Hammerling, and B. Dupont. 1987. Description of the reference panel of B-limphoblastoid cell lines for factors of the HLA system: The B-cell line panel designed for the Tenth International Histocompatibility Workshop. In *Immunobiology of HLA. Histocompatibility testing 1987* (ed. E. Dupont) pp. 11–19, Springer-Verlag, New York, NY.

Zhong, N., W. Yang, C. Dobkin, and W.T. Brown. 1995. Fragile X gene instability: Anchoring AGGs linked microsatellites. *Am. J. Hum. Genet.* 57: 351–361.

*Received October 7, 1996; accepted in revised form April 4, 1997.*