Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 5312–5316, June 1993
Immunology

A highly polymorphic microsatellite in the class II *Eb* gene allows tracing of major histocompatibility complex evolution in mouse

(tandem repeat/recombination hotspot/standard haplotype/wild haplotype)

BRATIN K. SAHA*†, JACK J. SHIELDS‡, RAYMOND D. MILLER§, TED H. HANSEN§, AND DONALD C. SHREFFLER§

*Department of Pathology, Emory University School of Medicine, Atlanta, GA 30322; and Departments of ‡Pathology and §Genetics, Washington University School of Medicine, St. Louis, MO 63110

Contributed by Donald C. Shreffler, February 9, 1993

ABSTRACT **A hallmark of major histocompatibility complex (MHC) genes is their extraordinarily high level of polymorphism. Polymorphic residues on MHC molecules determine which peptide ligands they bind and present to effector T lymphocytes. Although the genetic mechanisms responsible for MHC polymorphism have been delineated, the timetable and the pathway of their diversification remain unclear. To trace MHC evolution, we have characterized a highly polymorphic microsatellite containing tandem repeats (TRs) of two tetranucleotide units, TGGA and GGCA, located at the 3′ end of the second intron in the class II *Eb* gene of mouse. On the basis of length as well as sequence variations, 11 TR alleles were defined in 55 inbred mouse strains, which included MHC recombinant haplotypes and haplotypes derived from different subspecies of mouse. In this extensive sampling, a striking concordance was observed between the serologically identified class II proteins and the associated TR alleles. Examination of several strains carrying the same MHC haplotypes as well as strains carrying recombinant MHC haplotypes indicates that TR alleles are extremely stable. These observations suggest that TR polymorphism predates the separation of various subspecies of mouse. On the basis of sequence divergence, a genealogical tree has been constructed to depict evolution of the different TR alleles. Finally, evidence is presented that suggests this microsatellite polymorphism is generated by slipped-strand mispairing during DNA replication.**

The murine major histocompatibility complex (*H-2*) represents a cluster of tightly linked genes encoding antigens of the class I, class II, and class III families (1). The class II family encompasses four functional genes, *Ab*, *Aa*, *Eb*, and *Ea*, which encode the corresponding polypeptides. The A_β and A_α polypeptides make up the heterodimeric I-A molecule, whereas the E_β and E_α polypeptides make up the heterodimeric I-E molecule. Both the I-A and I-E molecules are expressed as glycoproteins on the surface of specialized antigen-presenting cells such as macrophages, B cells, and dendritic cells. During an immune response, these molecules bind processed peptides from exogenous antigens and present them to $CD4^+$ T lymphocytes (2).

The class II (as well as the class I) family represents the most polymorphic genes known to date. The high degree of polymorphism has been positively selected and maintained through evolution so that there are over 100 alleles at this locus in mouse and man. This diversity of major histocompatibility complex (MHC) molecules results in their ability to bind a wide variety of peptide ligands and has direct functional relevance for immune responses. Although the genetic mechanism contributing to MHC polymorphism has been delineated, the relative time over which this diversification occurred is still perplexing. While some studies suggest that the generation of MHC alleles is an ongoing process (3, 4), others indicate that certain MHC alleles are shared by different species (5, 6). On the basis of the latter observations, it has been proposed that many of the MHC alleles represent ancient polymorphisms that have survived millions of years of diversification of the *Mus* species (transspecies evolution).

In recent years much attention has been drawn to repeat DNA sequences as a tool for studying polymorphism and evolution. Tandem repeats (TRs) of simple DNA sequences are scattered throughout the mammalian genomes (7). These sequences are usually present within an intron and upstream or downstream from a gene. The tandem repeats are often characterized by a high degree of polymorphism, which leads to a great variation in the length of repeat sequences (8). Since these repeat sequences tend to vary extensively between individuals (strains) within a species, they have been extremely useful for identification of individuals, as well as for linkage mapping in pedigrees (9). Depending on the length, these polymorphic sequences are referred to as minisatellites [also called variable number of tandem repeats (VNTRs)] or microsatellites [also called simple sequence-length polymorphisms (SSLPs)].

Several types of microsatellites have been observed in man, mouse, and other eukaryotes. The most prominent of these are the $(CA)_n$ [or $(GT)_n$] repeats, 50–100,000 copies of which are interspersed throughout the human genome (10, 11). The $(CA)_n$ repeats have already been found associated with a few hundred genes and are rapidly becoming a staple for linkage analysis (12). A long stretch of $(GT)_n$ repeat, along with shorter stretches of $(GA)_n$ and $(CT)_n$ repeats, has also been observed at the 5′ region of the class II *Eb* gene in mouse (13). However, the extent of polymorphism in these microsatellites is not known.

In this report, we describe a highly polymorphic microsatellite made up of two tetranucleotide blocks, $(TGGA)_x$ and $(GGCA)_y$, located at the 3′ end of the second intron in the *Eb* gene. We investigate the stability of this microsatellite, demonstrate its application in the characterization of *H-2* haplotypes and recombinants, discuss its implication in transspecies evolution, and speculate on the mechanism of its diversification.

MATERIALS AND METHODS

Mice. The A.DB (1R) strain was obtained from David Sachs (14) through the Hazelton Laboratories America, Vienna, VA. All other strains (54 altogether) were bred and maintained in the Department of Genetics, Washington University School of Medicine, St. Louis.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MHC, major histocompatibility complex; SSLP, simple sequence length polymorphism; TR, tandem repeat.

†To whom reprint requests should be addressed.

DNA. Genomic DNA was isolated from the liver either directly or after isolation of the nuclei following a standard protocol. High molecular weight DNA was spooled out from 66% (vol/vol) ethanol and dissolved in water.

PCR. The primers were A, 5′-CGACTGTAGAACCT-TAGCCTG-3′ (2161–2181); B, 5′-TGGAGCTGTCCTCCT-TGTAG-3′ (2281–2300); and C, 5′-GTGGACACAATTCCT-GTTTTC-3′ (2465–2485). Primers A and B were used for determining the length polymorphism, and primers A and C were used for determining the allele sequence. The numbers in parentheses correspond to the nucleotide positions in figure 3 of ref. 15. The amplification was carried out for 35 cycles with AmpliTaq DNA polymerase (Perkin–Elmer/Cetus), following the supplier's protocol. Each cycle consisted of 1 min at 94°C, 2 min at 58°C, and 2 min at 72°C. DNA was heated at 95°C for 5 min at the beginning and at 72°C for 7 min at the end of cycles.

SSLP. DNA was labeled during amplification either internally (Fig. 2) with [^{32}P]dCTP or at the 5′ end (Fig. 4) with a 5′ ^{32}P-labeled primer. After diluting in 50% (vol/vol) formamide and heating at 95°C for 5 min, the length of the amplified DNA was determined by electrophoresis through a sequencing gel. Autoradiography was for 1–3 days.

DNA Sequencing. To minimize error due to *Taq* polymerase-induced misincorporation of nucleotides, DNA was sequenced directly (without cloning) after PCR amplification utilizing the *fmol* DNA sequencing system (Promega), and deoxyadenosine 5′-[γ-[^{35}S]thio]triphosphate. Sequence of only the variable portion of DNA is shown in Fig. 3. The alleles are named after the length of the DNA sequence presented, which is always 42 nt shorter than the DNA sequence amplified. Sequences for the haplotypes *b*, *d*, and *k*; *f* and *q*; and *s* were taken from refs. 15, 16, and 17, respectively.

RESULTS

Identification of TR Alleles. A tandem repeat of a tetranucleotide, GGCA, was observed at the 3′ end of second intron in the *Eb* gene of murine MHC (15). Interestingly, the number of repeats varied in four of the six *H-2* haplotypes sequenced (15–17). On further examination, we noticed another tetranucleotide repeat, TGGA, immediately 5′ to the GGCA repeat (Fig. 1). The number of each repeat varied considerably from haplotype to haplotype, and together these two repeats represented a highly polymorphic microsatellite. Potentially, many of the *H-2* haplotypes could be identified on the basis of length variation of this microsatellite.

We set up an assay based on PCR amplification of a segment of DNA encompassing the repeat sequences, followed by sizing the DNA fragments by urea/PAGE. In the 12 standard haplotypes examined (Table 1), seven length variations were detected (Fig. 2 *Left*). The haplotypes *b*, *k*, *r*, and *v* had the same length, as did *f* and *s*, and *u* and *z*, whereas *d*, *q*, *j*, and *p* each had a unique length. The *p* haplotype had the shortest sequence (50 nt), and the *u* and *z* haplotypes had the longest sequence (106 nt).

We then examined 19 wild-derived haplotypes (Table 1) and detected seven length variations as well, six of which were identical to those found in the standard haplotypes (Fig. 2 *Right*). TR 58 was not observed in the wild haplotypes,

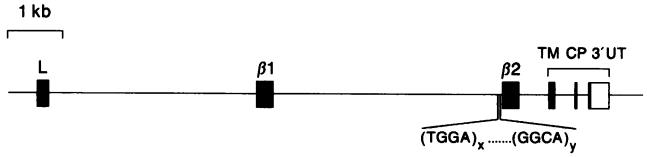

FIG. 1. Structure of the *Eb* gene indicating the position of the TR sequence. L, leader; TM, transmembrane; CP, cytoplasmic; 3′UT, 3′ untranslated sequences.

Table 1. *Eb* TR alleles in the standard and wild-derived strains

Strain*	*Eb*† allele	TR allele
FAIYUM-3	*w206*‡	106a
B10.NZW	*u* (z)	106b
B10.PL	*u*	106b
B10.D2	*d*	98
STU	*d*§	98
B10.MOL/K	*w*?	98
B10.LIB55(A)	*w13*	93a
B10.STA39(A)	*w13*	93b
B10.Q	*q*	74a
B10.WR7(A)	*w7*	74b
B10.M	*f*	70
B10.S	*s*	70
B10.DDD	*s*¶	70
B10	*b*	66
B10.BR	*k*	66
B10.RIII	*r*	66
B10.SM	*v*	66
B10.SNA70(A)	*k*	66
B10.W10LT(A)	*k*¶	66
B10.KPB128(A)	*s*	66
B10.GAA37(A)	*s*	66
B10.KPA42(A)	*v*	66
B10.BUA1(A)	*w16*	66
B10.BUA19(A)	*w16*	66
B10.MOL1	*k*?	66
B10.WB	*j*	58
B10.P	*p*	50
B10.CHR51(A)	*p*	50
B10.STC77(A)	*p*	50
C3H.W3(A)	*w*?	50
B10.BAC/K	*w*?	50

*The "A" in parentheses indicates an Ann Arbor strain.
†Data from ref. 18, except where indicated otherwise.
‡Data from ref. 19.
§Data from ref. 20.
¶Data from ref. 21.

whereas TR 93 was lacking in the standard haplotypes. Except for TR 93, each length variation differed from the other by a multiple of 4 nt, suggesting loss or gain of one or more of the tetranucleotide repeats.

To determine if the length variations represented any sequence heterogeneity, we amplified DNA from 25 of the 31 strains by PCR and sequenced them directly. The remaining six strains had already been sequenced (see *Materials and Methods*). Only three of the eight length variations showed any sequence differences (Fig. 3). These included TR 106 (FAIYUM-3 and B10.NZW or B10.PL), TR 93 (B10.LIB55 and B10.STA39), and TR 74 (B10.Q and B10.WR7). The two TR 106 alleles, as well as the two TR 74 alleles, demonstrated a considerable variation in the number of tetranucleotide repeats. The two TR 93 alleles, on the other hand, had swapped one repeat unit for the other. The largest group (composed of four standard and eight wild-derived haplotypes), which represented the length variation 66, all yielded an identical nucleotide sequence. Likewise, the standard haplotype *p* and four other wild-derived haplotypes all shared the same sequence (allele 50). Altogether 11 TR alleles were identified in the 31 strains examined (Fig. 3). The 106a and 106b, 93a and 93b, and 74a and 74b alleles can also be discriminated by single-stranded conformation polymorphism (SSCP) gel analysis, which circumvents the more time-consuming DNA sequencing (data not shown).

The 12 standard haplotypes represented seven alleles, and the 19 wild-derived haplotypes represented eight alleles. Three alleles (106b, 74a, and 58) were present only in the standard haplotypes, whereas four (106a, 93a, 93b, and 74b)

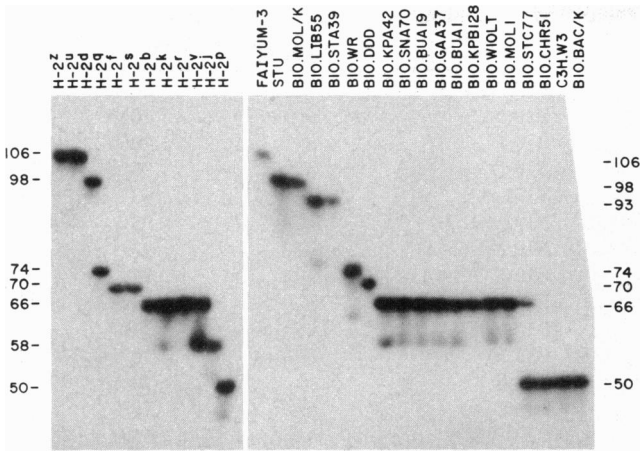

FIG. 2. Length variation of the microsatellite in the standard and wild-derived haplotypes. Numbers on each side represent the number of nucleotides as determined by sequencing (see Fig. 3). The actual size of the amplified DNA fragments is 42 nt higher (e.g., the length of the *H-2*^{z} allele is 148 nt and that of the *H-2*^{p} allele is 92 nt). Note that the *H-2*^{v} and B10.STC77 lanes were each contaminated with DNA from the adjacent lanes, *H-2*^{j} and B10.MOL1, respectively. The strains B10.NZW, B10.PL, B10.D2, B10.Q, B10.M, B10.S, B10, B10.BR, B10.RIII, B10.SM, B10.WB, and B10.P were used as representatives for the haplotypes *z, u, d, q, f, s, b, k, r, v, j,* and *p*, respectively.

were exclusive in the wild haplotypes; the remaining four (98, 70, 66, and 50) were found in both categories. The 13 Ann Arbor strains represented a sample of an extended natural population (not a local deme), and five TR alleles were detected in this group (Table 1).

The overall number of TGGA repeats varied from 3 to 13. Nine of the 12 standard haplotypes and 16 of the 19 wild-derived haplotypes had four or fewer TGGA repeats. The standard haplotypes *j* and *u* or *z*, along with the wild-derived strains B10.WR7, B10.LIB55, and B10.STA39, all had a substantially larger number, 7–13, of TGGA repeats. Similarly, the number of GGCA repeats varied from 3 to 17. Unlike TGGA, the GGCA repeats varied considerably throughout the standard, as well as the wild-derived, haplotypes (Fig. 3). The estimate of heterozygosity for the entire population is 80%, and the estimate is comparable between the standard and wild-derived strains (81% vs. 75%). The overall polymorphism information content (PIC; ref. 22) value is 0.78.

Stability of TR Alleles. Because the number of tetranucleotide repeats varied considerably from strain to strain, it was

106a	GATCC ATTC	///	(TGGA)2	TAGA	TGGA	GGTA	////	////	(GGCA)16	TGCA GGCA GCCTA
106b	..... .G..	///	(TGGA)13	////	////	////	////	////	(GGCA)10	//// //// .....
98	..... .G..	///	(TGGA)2	....	....	....	(GGCA)4	TGCA	(GGCA)9	....
93a	..... .G..	TGA	(TGGA)8	////	////	////	////	////	(GGCA)11	//// //// .....
93b	..... .G..	...	(TGGA)7	////	////	////	////	////	(GGCA)12	//// //// .....
74a	....	///	(TGGA)2	....	....	....	////	////	(GGCA)8	.... .A.. .....
74b	..... .G..	///	(TGGA)11	////	////	////	////	////	(GGCA)4	//// //// .....
70	....	///	(TGGA)2	....	....	....	////	////	(GGCA)7	.... .A.. .....
66	..... .G..	///	(TGGA)2	....	....	....	////	////	(GGCA)6	....
58	..... .G..	///	(TGGA)8	////	////	////	////	////	(GGCA)3	//// //// .....
50	..... .G..	///	(TGGA)4	////	////	////	////	////	(GGCA)5	//// //// .....

FIG. 3. Sequence of the TR alleles in the standard and wild-derived haplotypes. The alleles have been ordered according to decreasing length and aligned to maximize the number of TGGA and GGCA repeats. The "a" and "b" refer to alleles that cannot be separated by length variations (see Fig. 2). A period indicates an identity with the nucleotide above it, whereas a slash indicates the absence of the nucleotide below or above it.

of interest to determine if the TR alleles carried by different strains with the same *Eb* region are identical or not. We analyzed eight strains of *Eb*^{k} [B10.A (2R), B10.BR, B10.AL, B10.A, C3H, A.TL, AKM, and A/Sn] and five strains of *Eb*^{d} (B10.OL, B10.D2, DBA/2, B10.NZB, and BALB/c). Some of these strains were established independently while others were derivatives thereof. Each of the *Eb*^{k}, as well as each of the *Eb*^{d}, strains showed an identical length TR allele (results not shown). Further analysis revealed that all the *Eb*^{k} strains also had an identical sequence in their TR alleles (results not shown). It is, therefore, likely that the TR alleles are also identical in other strains carrying the same *Eb* region.

TR Alleles in *Eb* Recombinants. The high degree of polymorphism and the stability of the TR alleles suggest that these sequences can be used as a genetic marker to identify the *Eb* alleles. Several intra-I region recombinants have been shown to have crossed over within a narrow segment of DNA in the *Eb* intron 2 (referred to as the *Eb* hotspot), just 5′ to the TR sequence (15, 17, 23). We analyzed some of these as well as several new recombinants for the TR alleles (Table 2). Fig. 4 shows a urea/PAGE analysis of the PCR-amplified DNA from these recombinants along with their parental strains. The recombinant B10.GD had been analyzed before by DNA sequencing (15). A.DB(1R), a recombinant derived independently from the same parental alleles, had presumably crossed over in a direction opposite to B10.GD (14). Fig. 4 shows that B10.GD carries the TR allele of the *b* haplotype, whereas A.DB(1R) carries the TR allele of the *d* haplotype, thus marking the 3′ boundary and confirming the reciprocal orientation of the crossover. The 3′ boundary and the direction of crossover of the other recombinants, such as B10.TFR5, B10.TBFR6, B10.QSR2, and B10.THR2 are also apparent from this analysis. More recently, we have mapped the crossover site in B10.THR2 to a 1-kb stretch of DNA just 5′ to the TR allele (B.K.S., unpublished results). Several of the *k* and *s* recombinants (Table 2) have been reported earlier (17); only two examples are shown here (Fig. 4). In each instance the TR allele in the recombinant was identical in length as well as in sequence (data not shown) to one or the other parental strain, which rules out TR sequence as the site of recombination.

DISCUSSION

To the best of our knowledge, this is the first description of a compound microsatellite containing TRs of two tetranucleotide blocks, TGGA and GGCA. Weller *et al.* (24) noted a long tandem repeat of $(TGGA)_{165}$ located 1 kb upstream from the human myoglobin gene with a sequence divergence of about 20%. An even longer repeat, $(TGGA)_{249}$, with about the same level of sequence divergence and at the same relative location was observed by Boylan *et al.* (25) in the myelin

Table 2. *Eb* TR alleles in the intra-I region recombinants

Eb recombinant	Parental *Eb* alleles	*Eb* TR allele(s)
B10.GD	*d/b*	*b*
A.DB(1R)	*b/d*	*d*
B10.TFR5	*f/k*	*k*
B10.TBFR6	*k/f*	*f*
B10.THR2	*s/k/b*	*k, b*
B10.QSR2	*s/q*	*q*
B10.ASR7	*k/s*	*s*
B10.ASR11	*s/k*	*k*
B10.ASR1	*s/k*	*k*
B10.ASR12	*s/k*	*k*
B10.BASR1	*k/s*	*s*
B10.S(8R)	*k/s*	*s*
B10.S(9R)	*s/k*	*k*
B10.HTT	*s/k*	*k*

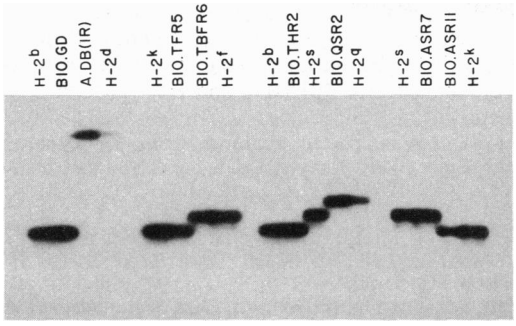

FIG. 4. TR alleles in the *Eb* recombinants and their parents. The parents are identified by their *H-2* alleles. The *H-2*^{d} lane was inadvertently underloaded (see Fig. 2 for a better visualization of *H-2*^{d} with respect to *H-2*^{b}).

basic protein gene. Although the extent of polymorphism in the myoglobin microsatellite is not known, the myelin basic protein microsatellite has been shown to be highly polymorphic, and 10 alleles have been identified. The microsatellite, $(TGGA)_x$ $(GGCA)_y$, described in this report is also highly polymorphic, and 11 alleles (referred to as TRs) have been detected thus far (Table 1 and Fig. 3).

TR Alleles Can Predict the *H-2* Serology. Our analysis of microsatellites in the standard *H-2* haplotypes defined seven TR alleles associated with *Eb*. To extend this analysis, several different mouse strains with wild-derived haplotypes were also studied. Importantly, a high percentage of wild-derived *H-2* haplotypes have been found to carry MHC alleles serologically identical to those found in laboratory strains. For example, Klein and colleagues (26) trapped mice in granaries at different locations around Ann Arbor, Michigan, in order to sample a wild-derived natural population. In a comprehensive serological analysis, the authors noted that several of these wild-derived *H-2* haplotypes carried class I and class II alleles whose products were indistinguishable from those in laboratory strains. However, serological typing of the *Eb* allele is more difficult than typing of the *Ab* or *Aa* allele. First, several of the mouse strains carry an *Eb/Ea* null allele so that there is no cell surface expression of the $E_\beta E_\alpha$ heterodimer, prohibiting any serology of the *Eb* or *Ea* polypeptide. Second, the I-E molecules appear serologically less complex than the corresponding I-A molecules. Therefore, the origin of *Eb* alleles in the wild-derived strains is frequently inferred from typing of the more polymorphic *Ab* and *Aa* alleles as well as their I-E phenotype (Table 1). In spite of these caveats, there is a remarkable correspondence between the assigned *Eb* alleles and the TR alleles identified here. Indeed, of the 10 wild-derived strains that have been assigned with a standard *Eb* allele, 8 are consistent with our designation of the TR alleles. The two exceptions are B10.KPB128 and B10.GAA37, which were assigned as *Eb*^{s}, and hence should carry the TR 70 allele. But our analysis indicates that both of these strains carry the TR allele 66. This could have resulted from a meiotic crossover between an *Eb*^{s} strain and one carrying a TR 66 allele (see below). Further studies including DNA sequencing are necessary to confirm this notion. In any case, the strong association between most of the TR alleles and the corresponding *Eb* alleles in both the standard and wild haplotypes indicates that these two markers are in linkage disequilibrium. A similar observation was made by Rieb *et al.* (27) in the HLA system, where the authors noticed a strong correlation between the sequence variability of the dinucleotide repeats, $(GT)_m(GA)_n$ (located in the second intron of the DRB gene, the human homologue of the mouse *Eb* gene) and serologic specificities of the *HLA-DRB* alleles.

TR Alleles Are Stably Inherited. Several lines of evidence suggest that the TR alleles are stably inherited. First, eight different strains containing the *Eb*^{k} allele, four of which were derived independently and included standard laboratory strains as well as wild-derived MHC haplotypes, were found to have an identical TR allele. The same was true for five different strains containing the *Eb*^{d} allele, although some of these strains had separated 100–200 generations ago. Second, analysis of over 50 independently derived *Eb* recombinants showed no alteration of the parental TR alleles although the crossover, in most instances, had occurred just 5′ to the TR sequence. Third, the TR alleles are shared not only within the same species but also across species as well. Thus the same TR alleles observed in *Mus domesticus* are also found in *Mus molossinus* (MOL/K and MOL1) and *Mus bactrianus* (BAC/K) (Table 1 and Fig. 2). Also several of the wild-derived strains examined in this report originated in different parts of the world. Thus MOL/K and MOL1 were originally from Japan, STU and DDD were from Germany, FAIYUM-3 was from Egypt, and the rest were derived from mice trapped in the Ann Arbor, Michigan, area. Yet these strains carried TR alleles that were either identical or similar to one other. Finally, as discussed earlier, there is a strong correspondence between the TR alleles and the *Eb* serology. Together these observations lead to the conclusion that TR alleles had evolved before the speciation/subspeciation of the mouse strains, probably 1–3 million years ago (28), and have since been maintained stably in the population. On the basis of serological study of two *Ab* alleles in several species of mice and rats, Figueroa *et al.* (29) also concluded that many of the MHC polymorphisms predated speciation.

TR Alleles Can Be Used to Trace Evolution. Nucleotide sequence of the TR alleles (Fig. 3) allows us to make certain inferences as to the origin and descent of these alleles. The ancestral TR allele had a simple structure of perfect repeats, $(TGGA)_x(GGCA)_y$. The number of each repeat varied through evolution due to unequal meiotic exchanges (30, 31) or slipped-strand mispairing during DNA replication (32, 33), creating additional alleles. These are referred to as group I alleles (Fig. 5*A*) and represent two strains from the standard haplotypes and seven from the wild-derived haplotypes. Eventually mutations started accumulating within the repeats introducing sequence heterogeneity and yielding newer alleles. Thus group IIA, representing the two wild-derived strains B10.LIB55 and B10.STA39, could have evolved by a simple point deletion of the guanine nucleotide (dG) from a TGGA repeat (Fig. 5*B*). It is intriguing to speculate that both these strains descended from a common ancestor by a balanced slipped-strand mispairing during replication of DNA (Fig. 5*C*). This represents a unique case where the two strains have an identical serologic specificity (both *w13*, Table 1) and yet carry a somewhat different TR allele. This would suggest that these two strains have diverged from each other in a more recent time and represent the latest members of TR alleles in the evolutionary tree. Group IIB, the largest group, includes 13 members, 4 from the standard and 9 from the wild-derived haplotypes, and could have been derived from group I by a series of point mutations (G to A, G to T or G to C, and C to T) involving both the repeats. Group IIIA is generated from group IIB by a second G to A substitution, whereas group IIIB is generated from group IIB by a second G to T substitution; both substitutions occurred in the GGCA repeat. FAIYUM-3 (group IIB) and group IIIA members incurred an additional G to T substitution in the 5′ anchor sequence.

On the basis of sequence analysis of the second exon of *Ab* alleles, Wakeland *et al.* (34) defined seven evolutionary groups among 8 standard and 18 wild haplotypes. Where direct comparisons can be made, some of the groupings using *Ab* sequences (34) and *Eb* TR sequences (Fig. 5) are identical. However, there are also notable differences. Some of the differences could be attributed to the "hotspot" of recombination separating these two loci. It is also noteworthy that

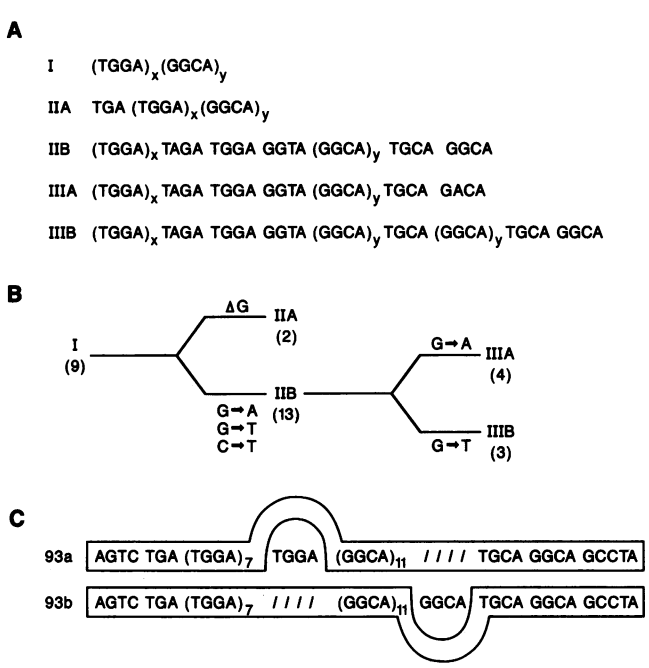

FIG. 5. (*A*) Five evolutionary groups of the *Eb* TR alleles. Group I, B10.PL and B10.NZB (both 106b), B10.WR7 (74b), B10.WB (58), and B10.P, B10.STC77, B10.CHR51, C3H.W3, and B10.BAC/K (all 50); group IIA, B10.LIB55(93a) and B10.STA39 (93b); group IIB, FAIYUM-3 (106a) and B10, B10.BR, B10.RIII, B10.SM, B10.W10.LT, B10.KPA42, B10.SNA70, B10.MOL1, B10.BUA1, B10.KPB128, B10.GAA37, and B10.BUA19 (all 66); group IIIA, B10.Q (74a) and B10.M, B10.S, and B10.DDD (all 70); and group IIIB, B10.D2, STU, and B10.MOL/K (all 98). The "x" and "y" refer to the number of respective repeats and are different for different alleles. (*B*) Proposed pathway for evolution of the different groups. The number in parentheses indicates the number of strains in that group. (*C*) Schematic showing comparison and proposed derivation of the 93a and 93b alleles.

tracing evolution using sequence of MHC gene exons may be obfuscated by their participation in intragenic exchanges or gene conversion-like events. Thus the inclusion of stable sequences, such as the TR allele, in studying evolution offers a unique approach to tracing genealogical relationship.

TR Alleles Are Probably Created by Slipped-Strand Mispairing. Microsatellites are often located close to a hotspot of recombination (17, 35, 36). This led to the speculation that unequal crossing-over during a meiotic recombination is the principal mechanism for generation of length hypervariability of the microsatellite (30). However the evidence is rather circumstantial and does not prove that the repeat sequences are themselves involved in the crossover. Further, there is no correlation between an increased number of repeats and an increased rate of recombination. Thus mouse genome, despite having a greater number of $(GT)_n$ repeats, undergoes significantly less recombination as reflected by fewer centimorgans (cM) in the genetic map of mouse (1600 cM) compared to human (2600 cM) (37). Also, as mentioned before, analysis of some 50 *Eb* recombinants indicated that crossovers, although always adjacent to, were never within the TR microsatellite (15, 17, 23). Wahls *et al.* (38), utilizing cultured human cells and constructs of the consensus microsatellite sequence, showed a 13.5-fold stimulation of homologous recombination. Thus it is more likely that repeat sequences somehow instigate and/or enhance homologous recombination. Morral *et al.* (39) showed by linkage analysis that there was no exchange of flanking sequences around a microsatellite in several CFTR recombinants, indicating that meiotic (or mitotic) recombination was less likely to be a cause for the polymorphism of this microsatellite. Addition or deletion of one or more repeat units, on the other hand, is more likely to be due to slipped-strand mispairing during DNA replication (32) as evidenced in B10.LIB55 and B10.STA39.

This work was supported by funds from the Department of Pathology (Emory University) (B.K.S.) and National Institutes of Health Grants AI-19687 (T.H.H.) and AI-12734 (D.C.S.).

1. Steinmetz, M., Minard, K., Horvath, S., McNicholas, J., Frelinger, J., Wake, C., Long, E., Mach, B. & Hood, L. (1982) *Nature (London)* **300**, 35–42.
2. Buus, S., Sette, A. & Grey, H. M. (1987) *Immunol. Rev.* **98**, 115–141.
3. McIntyre, K. & Seidman, J. G. (1984) *Nature (London)* **308**, 551–553.
4. Nathenson, S. G., Geliebter, J., Pfaffenbach, G. M. & Zeff, R. A. (1988) *Annu. Rev. Immunol.* **4**, 471–502.
5. Arden, B. & Klein, J. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2342–2346.
6. Sagai, T., Sakaizumi, M., Miyashita, N., Bonhome, F., Petras, M. L., Nielson, J. T., Shiroishi, T. & Moriwaki, K. (1989) *Immunogenetics* **30**, 89–98.
7. Hamada, H., Petrino, M. G. & Kakanuga, T. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 6465–6469.
8. Tautz, D. (1989) *Nucleic Acids Res.* **17**, 6463–6471.
9. Jeffreys, A. J., Neumann, R. & Wilson, V. (1990) *Cell* **60**, 473–485.
10. Weber, J. L. & May, P. E. (1989) *Am. J. Hum. Genet.* **44**, 388–396.
11. Litt, M. & Luty, J. A. (1989) *Am. J. Hum. Genet.* **44**, 397–401.
12. Weber, J. L. (1990) *Genomics* **7**, 524–530.
13. Gillies, S. D., Folsom, V. & Tonegawa, S. (1984) *Nature (London)* **310**, 594–597.
14. Epstein, S. L., Arn, S. & Sachs, D. H. (1990) *Immunogenetics* **32**, 142–145.
15. Kobori, J. A., Strauss, E., Minard, K. & Hood, L. (1986) *Science* **234**, 173–179.
16. Begovitch, A. B., Thanh, H. V. & Jones, P. P. (1990) *J. Immunol.* **144**, 1957–1964.
17. Padgett, K. A., Shreffler, D. C. & Saha, B. K. (1991) *J. Immunol.* **147**, 2764–2770.
18. Klein, J., Figueroa, F. & David, C. S. (1983) *Immunogenetics* **17**, 553–596.
19. Wakeland, E. K., Price-LaFace, M., Henson, V. & Peck, A. B. (1987) *Immunogenetics* **26**, 115–119.
20. Wakeland, E. K. & Darby, B. R. (1983) *J. Immunol.* **131**, 3052–3057.
21. Hauptfeld-Dolejsek, V. & Shreffler, D. C. (1989) *Immunogenetics* **33**, 132–140.
22. Botstein, D., White, R. L., Skolnick, M. & Davis, R. W. (1980) *Am. J. Hum. Genet.* **32**, 314–331.
23. Zimmerer, E. J. & Passmore, H. C. (1991) *Immunogenetics* **33**, 132–140.
24. Weller, P., Jeffreys, A. J., Wilson, V. & Blanchetot, A. (1984) *EMBO J.* **3**, 439–446.
25. Boylan, K. B., Ayres, T. M., Popko, B., Takahashi, N., Hood, L. E. & Prusiner, S. B. (1990) *Genomics* **6**, 16–22.
26. Klein, J. & Figueroa, F. (1981) *Immunol. Rev.* **60**, 23–57.
27. Rieb, O., Kammerbaur, C., Roewer, L., Steimle, V., Andreas, A., Albert, E., Nagai, T. & Epplen, J. T. (1990) *Immunogenetics* **32**, 110–116.
28. Potter, M., Nadeau, J. H. & Cancro, M. P. (1986) *Curr. Top. Microbiol. Immunol.* **127**, 12–19.
29. Figueroa, F., Gunther, E. & Klein, J. (1988) *Nature (London)* **335**, 265–267.
30. Jeffreys, A. J., Wilson, V. & Thien, S. L. (1985) *Nature (London)* **314**, 67–73.
31. Coggins, L. W. & O'Prey, M. (1989) *Nucleic Acids Res.* **17**, 7417–7426.
32. Levinson, G. & Gutman, G. A. (1987) *Mol. Biol. Evol.* **4**, 203–221.
33. Wolf, R. K., Plaetke, R., Jeffreys, A. J. & White, R. (1989) *Genomics* **5**, 382–384.
34. Wakeland, E. K., Boehme, S., She, J. X., Lu, C.-C., McIndoe, R. A., Cheng, I., Ye, Y. & Potts, W. K. (1990) *Immunol. Res.* **9**, 115–122.
35. Treco, D. & Arnheim, N. (1986) *Mol. Cell. Biol.* **6**, 3934–3947.
36. Pardue, M. L., Lowenhaupt, K., Rich, A. & Nordheim, A. (1987) *EMBO J.* **6**, 1781–1789.
37. Stallings, R. L., Ford, A. F., Nelson, D., Torney, D. C., Hildebrand, C. E. & Moyzis, R. K. (1991) *Genomics* **10**, 807–815.
38. Wahls, W. P., Wallace, L. J. & Moore, P. D. (1990) *Cell* **60**, 95–103.
39. Morral, N., Nunes, V., Casals, T. & Estivill, X. (1991) *Genomics* **10**, 692–698.