Gene Conversion in the Evolution of Both the *H-2* and *Qa* Class I Genes of the Murine Major Histocompatibility Complex

Mary Kuhner,* Suzanne Watts,† William Klitz,* Glenys Thomson* and Robert S. Goodenow†

**Department of Integrative Biology and Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, and †Cell Biology Department, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030*

Manuscript received January 27, 1990
Accepted for publication September 4, 1990

ABSTRACT

In order to better understand the role of gene conversion in the evolution of the class I gene family of the major histocompatibility complex (MHC), we have used a computer algorithm to detect clustered sequence similarities among 24 class I DNA sequences from the *H-2*, *Qa*, and *Tla* regions of the murine MHC. Thirty-four statistically significant clusters were detected; individual analysis of the clusters suggested at least 25 past gene conversion or recombination events. These clusters are comparable in size to the conversions observed in the spontaneously occurring *H-2K*^{bm} and *H-2K*^{km2} mutations, and are distributed throughout all exons of the class I gene. Thus, gene conversion does not appear to be restricted to the regions of the class I gene encoding their antigen-presentation function. Moreover, both the highly polymorphic *H-2* loci and the relatively monomorphic *Qa* and *Tla* loci appear to have participated as donors and recipients in conversion events. If gene conversion is not limited to the highly polymorphic loci of the MHC, then another factor, presumably natural selection, must be responsible for maintaining the observed differences in level of variation.

THE class I genes of the murine major histocompatibility complex (MHC), located on chromosome *17*, can be divided into three distinct regions: *H-2*, *Qa*, and *Tla* (HOOD, STEINMETZ and MALISSEN 1983). Figure 1 shows the arrangement of these regions in several mouse strains. The *H-2* or "classical" loci *K*, *D* and *L* encode the transplantation antigens, comparable to *HLA-A* and *-B* in humans. These loci are extraordinarily polymorphic, with amino acid variability concentrated primarily in the antigen recognition site (ARS), a cleft formed by two α-helices lying across a β-pleated sheet on the top of the molecule (BJORKMAN *et al.* 1987a,b). The T cell receptor interacts specifically with residues on the α-helices, as well as with peptide bound in the cleft, in MHC-restricted cytotoxic T cell responses (BJORKMAN *et al.* 1987b). The "nonclassical" loci of the *Qa* and *Tla* regions, which have the same general protein and gene structure as the H-2 molecules (CHEN *et al.* 1987; ROBINSON 1987; WEISS 1987), display much less variability–few alleles can be distinguished serologically or biochemically–and do not appear to function as restriction elements (CHEN *et al.* 1987; ROBINSON 1987).

Elucidating the history of *H-2* polymorphism is an intriguing evolutionary problem. Analysis of spontaneous mutants including *H-2K*^{km2} and the *H-2K*^{bm} series has revealed one diversification mechanism in which small sequence transfers between loci, termed gene conversion, introduce multiple amino acid changes (NATHENSON *et al.* 1986; GELIEBTER and NATHENSON 1988; VOGEL *et al.* 1988). The mutations usually appeared in several littermates, reminiscent of mitotic recombination in the germ line (NATHENSON *et al.* 1986). It has been suggested, based on the structures of these mutants and the patchwork pattern of homology among naturally occurring *H-2* and *HLA* sequences (PEASE 1985), that gene conversion is a major mechanism for generating new class I alleles (NATHENSON *et al.* 1986; PEASE 1985). However, since other multigene families, such as hemoglobin, in which gene conversion also operates (COLLINS and WEISSMAN 1984), do not share the high class I polymorphism, forces other than conversion itself must be at work.

Natural selection for diversity and heterozygosity is also thought to promote the establishment and maintenance of histocompatibility antigen variability (KLITZ, THOMSON and BAUR 1986; KLEIN 1986). This theory is consistent with the observation that most of the polymorphism among *H-2* sequences consists of substitutions altering the amino acid sequence of the antigen recognition site (ARS) (HUGHES and NEI 1988). In contrast, the relative lack of polymorphism among *Qa* and *Tla* alleles suggests that they have evolved under different constraints than the *H-2* genes. Although their function remains unknown, possible roles in development or immune regulation have been suggested for these molecules based on their ontogenic- and tissue-specific distribution (FLAGHERTY *et al.* 1990; MORSE *et al.* 1990). Alternatively, the *Qa* and *Tla* loci may represent deteriorating class I genes, some of which encode nonfunc-

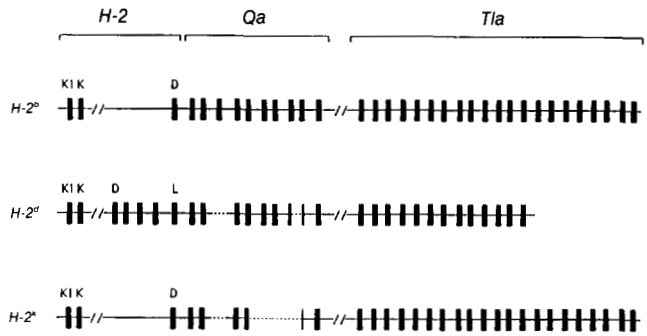

FIGURE 1.—Physical map of the class I loci. Schematic representation of the physical map of the class I genes from three inbred mouse strains, C57BL/10 ($H\text{-}2^b$), BALB/C ($H\text{-}2^d$) and C3H ($H\text{-}2^k$). Genes (boxes) within the *H-2* and *Qa* regions are positioned to show allelic relationships based on DNA sequence similarity. Dashed lines indicate space added to facilitate this positioning. Note that the locus named *L* on haplotype $H\text{-}2^d$ (and $H\text{-}2^q$) is apparently allelic to locus *D* of other haplotypes, while D^d (and D^q) represents a divergent duplication; the nomenclature predates the molecular map of this region. Boxes in the *Tla* region represent estimated numbers of genes only, not allelic relationships. Adapted from WATTS *et al.* 1989.

tional vestigial proteins (HUGHES and NEI 1989). Correspondingly, their near monomorphism may result from selection for a particular structure, or lack of the balancing selection which apparently promotes *H-2* variability. Since both *Qa* and *Tla* genes have been identified as potential donors for $H\text{-}2K^{bm}$ mutations (GELIEBTER and NATHENSON 1988), it has been proposed that the nonclassical loci may be evolutionarily important in that they provide a reservoir of sequence information for generating novel *H-2* alleles by gene conversion (BREGEGERE 1983).

We have used a statistical algorithm (STEPHENS 1985) to detect evidence of clustering indicative of past gene conversion events among 24 murine class I sequences. For each putative conversion event we have analyzed the sequences and gene regions involved, the apparent length of the event, and its effect on protein sequence in order to gain a clearer picture of the nature, extent, and evolutionary significance of genetic exchange in the evolution of the mouse MHC.

MATERIALS AND METHODS

DNA sequences: An alignment with references for all the *H-2* and *Qa* sequences used can be found in KUHNER and GOODENOW (1989); for this study we have added the *Tla* sequence *T13* (FISHER, HUNT and HOOD 1985). Only exon sequences were compared due to the difficulty of aligning introns among the more distantly related sequences. Future reports may be able to confirm or reject some of the events reported here by using intron data. Because of the remarkable sequence conservation between the $H\text{-}2K^q$ cDNA and genomic $H\text{-}2K^{w28}$ (MORITA *et al.* 1985), the latter was used to represent $H2\text{-}K^q$ in regions not covered by the truncated cDNA. As a test of the method, we also used human sequences *HLA-A1* (PARHAM *et al.* 1988), *A2.1* (HOLMES *et al.* 1987), *B8* (PARHAM *et al.* 1989), *B51* (ENNIS *et al.* 1990), *Cw1* (GUSSOW *et al.* 1987), and *Cw3* (SODOYER *et al.* 1984), and bovine sequences *BL3-6* and *BL3-7* (ENNIS, JACKSON and PARHAM 1988) for one comparison. Base pairs deleted or unknown in any sequence were omitted from all comparisons involving that sequence.

Statistical algorithm: A modified form of the method of STEPHENS (1985) was used. Briefly, each variable nucleotide site is treated as a division of the group of sequences under consideration into two subgroups. All sites which produce the same subgroups are analyzed together as a set, called a "partition." For example, the sites at which $H2\text{-}K^d$ and $Q5^k$ share one base pair, while all other sequences share a different base pair, constitute a partition. Non-random distribution–clustering–of the sites in a partition is taken as evidence for genetic exchange. If two or more sequences share substitutions because they have a common ancestor, the substitutions are expected to be randomly distributed across all variable sites in the molecule. If, on the other hand, substitutions are shared because of genetic exchange, they will be found predominantly in the region of exchange, forming a cluster.

Clustering was evaluated in two ways: comparison of the length spanned by the partition (d_0) to the length of the sequence, and comparison of the longest gap within the partition (g_0) to d_0. STEPHENS (1985) gives statistics for calculating the probabilities of d_0 and g_0 greater than or equal to those observed under the null hypothesis of random distribution of sites (P_{d_0} and P_{g_0}, respectively).

Two changes were made to adapt STEPHENS' method to the analysis of the class I data: (1) The distribution of variability across the class I gene is not random; the antigen presenting sites are hypervariable, while the more 3′ domains are highly conserved. To partially correct for this, all sites found to be invariant in a given comparison were excluded from that comparison. (2) Sites with more than two alternative nucleotides present a problem for STEPHENS' method. We chose to replace each such site with the nearest 2-alternative site of compatible form. For example, a 3-way division of A/B/C would be replaced by A/BC or AB/C or AC/B, the choice being made by searching along the sequence in both directions until a 2-way site of one of these types was found. If no similar 2-way site could be found, the 3-way site was omitted from analysis. This method allows comparison of large numbers of sequences simultaneously, and greatly increases the power of comparisons between divergent sequences (such as classical versus non-classical genes). However, it does introduce a bias toward identification of partitions as significant. To compensate for this and for the multiple comparisons made, we have reported only partitions with P_{d_0} or P_{g_0} less than 0.01. Computer simulations (unpublished results) suggest that this corresponds to a true significance level of less than 0.05. Furthermore, the majority of the clusters detected with this modification were also detected when sites with three or more bases were simply omitted from the calculations (results not shown).

In comparisons among large numbers of sequences, overlapping events and point mutations can obscure the detection of clusters. Therefore, in addition to the global comparison of all 24 sequences, we also compared *H-2* loci and *Qa* loci separately, and all comparisons were repeated with and without the most deviant loci, *T13* and $K1^k$. Most clusters were detected in more than one comparison, suggesting that the number of clusters detected was not being significantly inflated by multiple comparisons. Separate comparison of *H-2* and *Qa* loci also controls for the possibility that differences in the pattern of selection on these loci might create spurious evidence for gene conversion in the overall comparison. Table 1 shows the comparisons made.

TABLE 1

Subsets of sequences analyzed

Comparison	Abbr.	Seq.	Sites (%)	D	G	Sig
All mouse sequences	All	24	447 (74)	39	26	13
All except $K1^k$, *T13*	H2/Qa	22	414 (78)	46	29	17
H-2 including $K1^k$	H2/K1	13	262 (85)	36	22	7
H-2 except $K1^k$	H2	12	224 (84)	37	22	8
Qa and *T13*	Qa/Tla	11	390 (75)	30	20	12
Qa	Qa	10	350 (82)	31	27	15
Mouse, cow and human[a]	Cross	32	491 (66)	50	35	10

Seq. indicates number of sequences in comparison; sites, number of variable sites present (parenthetical value indicates percent of variable sites at which exactly two nucleotides were observed: remaining sites were equivalenced to two-way sites as described in MATERIALS AND METHODS). D indicates how many partitions were tested using P_{a0} (at least two sites required) and G how many were tested using P_{g0} (at least three sites required). Sig indicates how many partitions were significant at the 0.01 level for either or both statistics.

[a] All mouse sequences, plus human sequences *HLA-A1*, *A2.1*, *B8*, *B51*, *Cw1* and *Cw3*, and bovine sequences *BL3-6* and *BL3-7*. Analysis was of positions 1–888 only due to difficulty of aligning the most 3′ exons. This comparison was included to test the possibility that use of distantly related sequences would result in spurious SCPs, and is not analyzed in detail.

Many of the clusters identified by this algorithm are visible to the eye when demarcated on an alignment of the DNA sequences. However, several significant clusters would probably be missed, either because the informative nucleotide positions are widely spread, or because intervening mismatches obscure the clustered sequence similarity.

As pointed out by C. STEPHENS (personal communication) each cluster must be examined individually to determine whether genetic exchange is a plausible explanation for it, and if so what type of exchange could be involved: sequences can share clustered substitutions either because of genetic exchange at the location of the cluster, or because of exchange with an outside sequence which eliminates shared ancestral substitutions except in the region of the cluster.

Similarity trees: Trees of sequence similarity were constructed using program NJTREE, kindly provided by Li Jin, using the neighbor-joining distance algorithm (SAITOU and NEI 1987). *H-2* and *Qa* were considered separately; due to possible differences in evolutionary rate, we did not feel that constructing a tree for the entire data set would be meaningful. Trees were rooted on the midpoint between the two most divergent sequences.

RESULTS

A total of 34 significantly clustered partitions (SCPs) were detected using our adaptation of STEPHENS' algorithm in the six comparisons made (see Table 1 for a list of the comparisons). Each is described individually below; Table 2 presents an overview, and Figure 2 shows the distribution of the clusters on a map of the gene sequence.

As an aid to interpreting the clusters observed, we also constructed neighbor-joining trees of the sequences (Figure 3). We wish to emphasize that these trees probably do not represent the true tree of descent of the sequences due to unequal rates of evolution, and cannot represent the true relationships among sequences which have undergone genetic exchange. Rather, they should be considered convenient organizing devices which represent the sequence similarity of the alleles analyzed, and can help in the interpretation of observed clusters.

For each cluster, we have listed the alleles showing the apparently converted sequence, the base pairs involved (numbered as in the published alignment of these sequences (KUHNER and GOODENOW 1989), which counts exon positions only), and the comparisons in which this clustering was statistically significant. Level of significance (for the statistic P_{d_0} unless P_{g_0} is specified) is given by asterisks after the name of the comparison: *, $P < 0.01$; **, $P < 0.001$; ***, $P < 0.0001$; ****, $P < 0.00001$. For example, SCP #1 consists of clustering of $Q4^k$ unique sites at positions 1 and 2; it was detected in comparisons of *Qa* sequences with and without *Tla*; and in both of these comparisons, P_{d_0} was significant at the 0.01 level. Some of the SCPs detected included sites remote from the actual cluster (for example, partition #2d includes bases spanning an intron and far removed from the rest of the cluster, which are clearly not part of the cluster in the leader sequence); such sites are underlined. We also discuss possible conversion events which could have produced each cluster, the apparent length of the conversion, and its possible effect on the protein structure (unless specified otherwise, all clusters contained substitutions which changed the amino acid sequence).

When multiple lists of bases are given for a single SCP, there were substantial differences between the results from different subsets of sequences. For example, if the divergent *T13* sequence shares four of five sites involved in a clustering, comparisons including *T13* will report a four-site SCP, while comparisons excluding *T13* will report a five-site SCP, with correspondingly different probability values. Clusters identified as gaps in otherwise evenly distributed partitions were listed separately at the end, as in most cases it was difficult to assign these to events in a particular part of the gene.

The *T13* sequence is truncated at position 986, and therefore comparisons which involve it cannot detect any SCPs beyond this position. Similarly, the K^s se-

TABLE 2

Summary of significantly clustered partitions (SCPs)

Sequences	Location	Length	Interpretation
1. $Q4^k$	Leader		X to *Qa*
2. $Q4^{b,k}$,$Q7^{b,d}$,/$Q8^b$	Leader		*Qa* or X to *Qa*
3. $Q8^b$,$Q10^{b,k}$	α1	18	*Qa* to *Qa*
4. $D^{d,p}$,$K^{b,s}$	α1	32	*H-2* to *H-2*
5. $Q5^k$,*T13*	α1, α2		*Tla* to *Qa*
6. D^p,$K1^k$,$Q7^{b,d}$,$Q8^b$,$Q10^{b,k}$	α1	8	Complex
7. $K^{k,q}$	α1	45	X to *H-2*
8. K^d,$Q5^k$	α1	34	*Qa* to *H-2*
9. $D^{d,q}$,$K^{d,k,s}$,$Q5^k$,$Q7^{b,d}$,$Q8^b$	α1	16	Complex
10. K^b,$Q4^{b,k}$	α2	57	*Qa* to *H-2*
11. $Q4^{b,k}$,$Q5^k$,$Q10^{b,k}$	α2		Complex
12. D^p,$K1^k$			Artifact of algorithm
13. D^k,$K^{b,k,q,s}$	α2	101	*H-2* to *H-2*
14. K^d,L^d,$Q2^k$,$Q5^k$,$Q10^{b,k}$	α2	20	Complex
15. $Q2^k$,$Q4^{b,k}$,$Q7^{b,d}$,$Q8^b$	α2	61	*Qa* to *Qa*
16. $D^{b,d,q}$,$L^{d,q}$,$K^{b,q,s}$,*T13*	α2	21	Complex
17. $W7^{b,d}$,$Q8^b$	α2, α3		*Qa* or X to *Qa*
18. $K^{b,k,q,s}$	α3	214	X to *H-2*
19. D^d,$K^{b,k,q,s}$	α3	12	*H-2* to *H-2*
20. D^d,$Q5^k$,*T13*	α3, TM		*Qa* to *H-2*
21. $D^{d,p}$,$K1^k$,$Q5^k$	α3		Complex
22. D^d,$K^{d,k,q,s}$	TM	22	Complex
23. $Q1^k$,$Q2^k$,$Q5^k$	TM	22	*Qa* to *Qa*
24. $Q4^{b,k}$,$Q8^b$	TM, CYT		*Qa* to *Qa*
25. All *K* alleles	TM		Ambiguous
26. All *K*, $K1^k$,$Q1^k$	TM	15	*H-2* to *Qa*
27. $Q5^k$,$Q7^{b,d}$	CYT		*Qa* to *Qa*
28. $Q10^{b,k}$			Background
29. $Q7^{b,d}$			Background
30. $Q5^k$,$Q10^{b,k}$			Background
31. $Q5^k$			Background
32. $Q1^k$			Background
33. $Q2^k$			Background
34. $Q4^{b,k}$			Background

Length is maximum extent of identity between proposed donor and recipient, in base pairs. The abbreviation "X" refers to a hypothetical class I sequence not included in the present dataset. Clusters with the interpretation "Ambiguous" can be interpreted in more than one way, implicating different sets of sequences in the conversion; those with the interpretation "Complex" appear to require more than one conversion event for their explanation, while those with the interpretation "Background" appear to be the result of conversions elsewhere in the genes, rather than independent instances of conversion.

quence begins at position 11, and so *H-2* comparisons cannot detect SCPs in positions 1–10.

Analysis of individual SCPs: Analyses are as follows.

Leader (Exon 1: 1-74):

1. $Q4^k$ 1, 2 (*Qa/Tla**, *Qa**)

This allele shows two unique substitutions at the very beginning of the leader (upstream of the standard translation start). These substitutions and two apparent point mutations scattered across the $Q4^b$ sequence are the only differences between the *Q4* alleles, suggesting a possible conversion from an outside donor into $Q4^k$.

2a. $Q4^{b,k}$,$Q7^{b,d}$,$Q8^b$ 21, 38, 60, 62 (All**)
2b. 20, 21, 38, 60, 62 (*H2/Qa****)
2c. 6, 19, 21, 24, 38, 60, 62, 77 (*Qa/Tla*****)
2d. 6, 19, 20, 21, 24, 38, 60, 62, 77, 561, 937 (*Qa*:P_{g_0}**)

These alleles share a number of unique substitutions in the leader, while elsewhere *Q4* is more distinct from *Q7/Q8*. The *Qa* tree indicates that *Q4*, *Q7* and *Q8* are high similar (Figure 3D), and continues to do so even when the leader is omitted (data not shown); this SCP can be interpreted as an exchange between *Q4* and *Q7/Q8*, or as the result of exchanges elsewhere in the gene disrupting similarity between *Q4* and *Q7/Q8* (*i.e.*, SCPs #10 and #11). The similarity does not continue into intron 1 (data not shown), suggesting that position 77 (in the α1 domain) is not part of this event.

Alpha-1 (Exon 2: 75-344):

3. $Q8^b$,$Q10^{b,k}$ 142, 143, 144 (*Qa/Tla****, *Qa****)

The *Q8* locus clusters with *Q10* here rather than with its close relative *Q7*, suggesting a *Q10* to *Q8* transfer which changed ARS-homologous positions. This should be compared with the large number of leader positions at which *Q7* and *Q8* share substitu-

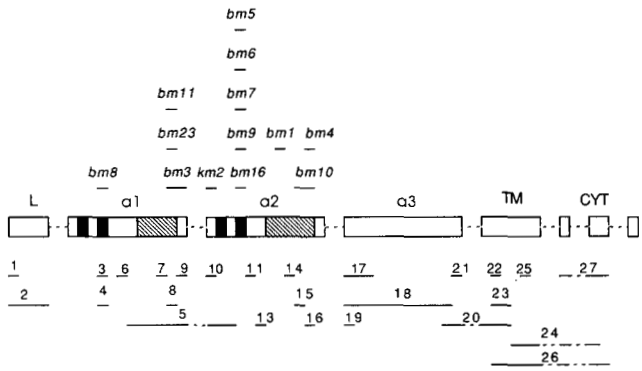

FIGURE 2.—Distribution of SCPs across the gene. A general schematic of the class I gene is shown; the more 3′ exons are not expressed in all loci. L, Translated leader sequence; TM, transmembrane domain; CYT, cytoplasmic domains. Black indicates the beta sheets and hatching the alpha helices which make up the antigen recognition site. The locations of SCPs reported in this paper and of the *H-2K*^{bm} (NATHENSON *et al.* 1986) and *H-2K*^{km} (VOGEL *et al.* 1988) mutants are shown. Note that introns are not drawn to scale (their length varies markedly between loci), and that the smaller SCPs have been enlarged to make them visible.

tions (SCP #2). SCPs #3 and #4 are illustrated in Figure 4.

4. $D^{d,p}$,$K^{b,s}$ 142, 144, 145 (*H2/K1***, *H2***)

Two related *H-2D* and two related *H-2K* alleles share a cluster of substitutions, suggesting a transfer from one locus to the other before divergence of these allele pairs. Amino acids in the ARS were affected.

5. $Q5^k$, *T13* 189, 200, 277, 321, 337, 340, 342, 409, 419 (All**)

Between positions 189 and 419 $Q5^k$ and *T13* share many unique substitutions, in addition to showing substitutions (not listed) common to only one or two other alleles. $Q5^k$ is otherwise a fairly typical *Qa* allele, suggesting transfer from *Tla* to *Qa*. However, there are also extensive mismatches between $Q5^k$ and *T13* in this region, suggesting either that the event is very old or that a *TLA* sequence distinct from *T13* served as donor to $Q5^k$. The event extends across intron 2, and affects ARS-homologous sequences in both α1 and α2.

6. D^p,$K1^k$,$Q7^{b,d}Q8^b$,$Q10^{b,k}$ 206, 207, 208 (*Qa/Tla****, *Qa****)

This clustering is significant in the comparison among *Qa* sequences, suggesting a transfer between *Qa* loci in addition to transfers involving D^p and $K1^k$. $Q5^k$ has a completely deviant sequence at these positions for which no donor can be found. Thus, this region, which is not part of the ARS, may retain traces of several conversion events.

7. $K^{k,q}$ 237, 240 (*H2/K1**)

Two substitutions are shared by the very similar K^k and K^q, suggesting a conversion from an outside donor–possibly a $Q5^k$-like sequence, as $Q5^k$ shares the substitution at 240 and an additional substitution at 244 which is present in K^k, K^q and K^b. Mismatches between K^k and K^q suggest an exchange length of 45 bp or less, but if point mutations have overlaid the event it could be much longer. It affects a non-ARS structural region.

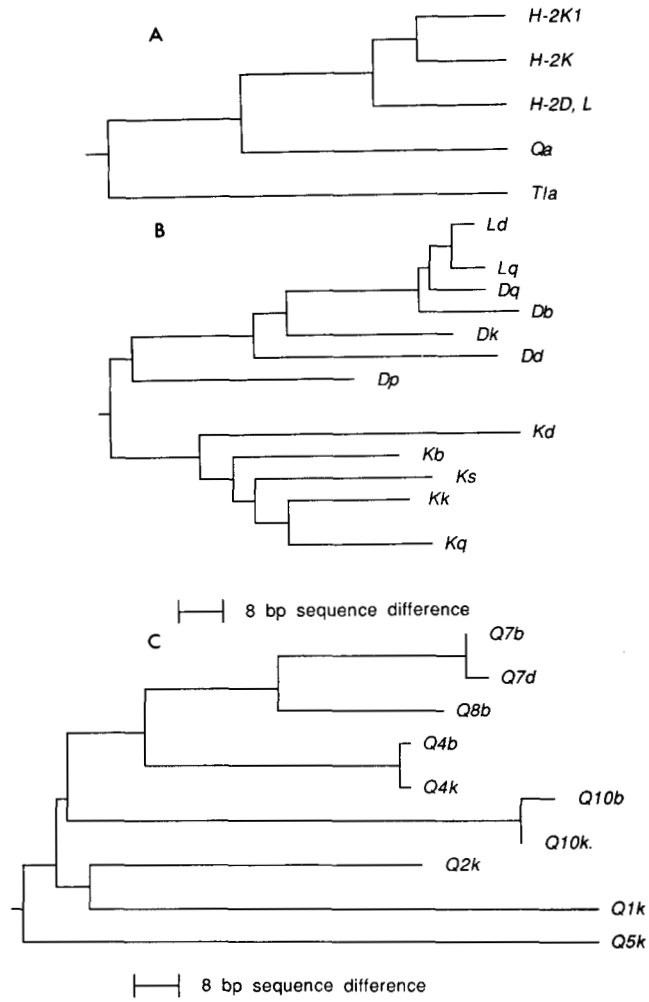

FIGURE 3.—Similarity relationships among sequences. (A) Schematic (not to scale) of the accepted relationships among the class I loci. (B) Neighbor-joining tree of *H-2* sequences. (C) Neighbor-joining tree of *Qa* sequences.

8. K^d,$Q5^k$ 257, 260, 278 (*H2/Qa**)

This region contains a cluster of similarities between K^d and $Q5^k$, but also four mismatches suggesting that the event is not recent. Sharing of bases between these sequences and *T13* (see SCP #5) suggests that the direction of transfer was $Q5^k$ to K^d. The event affects the ARS.

9a. $D^{d,q}K^{d,k,s}$,$Q5^k$,$Q7^{b,d}$,$Q8^b$ 314, 315 (All*, *Qa/Tla**)

9b. 314, 315, 319 (*H2/Qa***, *H2/K1***, *H2***)

9c. 314, 315, 319, 379 (*Qa**)

$D^{b,k,q}$, $L^{d,q}$	ctct
$K^{k,q}$	....
$Q1^k$,$Q2^k$,$Q4^{b,k}$,$Q5^k$,$Q7^{b,d}$	....
K^d	.g..
$K1^k$	.g..
T13	.g..
$D^{d,p}$	ggaa (SCP #4)
$K^{b,s}$	ggaa (SCP #4)
$Q10^{b,k}$	tat. (SCP #3)
$Q8^b$	tatc (SCP #3)

FIGURE 4.—Positions 142–148, illustrating SCPs #3 and #4.

Two disimilar *H-2D* alleles, three *H-2K* alleles, *Q5*, *Q7*, and *Q8* share a distinctive sequence. Within this pattern, there is especially strong similarity between K^d, *Q7* and *Q8*. In the cross-species comparison, at positions 314–315 the human allele *HLA-B51* shows the pattern of D^d, while other human and bovine alleles analyzed show the mouse consensus pattern ($P_{d_0} = 0.004$ in the cross-species comparison), suggesting either conservation of polymorphism at this position for extremely long periods of time, or else repeated mutation to the same two alternative forms. The large number of gene conversions which would be necessary to explain this SCP, and its presence in distantly-related species, suggest that convergent evolution may be contributing to its wide distribution, especially as this portion of the ARS is extremely immunogenic (the first class I sera developed recognized human variants in this region) and therefore a plausible target for strong selection. The complexity of this region is illustrated in Figure 5.

If the interpretation of SCP #5 as a conversion from *T13* to $Q5^k$ is correct, events in this region have evidently overlaid SCP #5 creating mismatches between *T13* and $Q5^k$.

Alpha (Exon 3: 345–620):

10a. K^{b},$Q4^{b,k}$ 355, 358, 363, 367 (*H2/Qa*****)
10b. K^b 355, 358, 363, 367, 379, 388, 397, 747, 936, 939 (*H2*: P_{d_0}*, P_{g_0}*)

Allele K^b shares a substantial run of sites with the *Q4* locus, suggesting a *Q4* to *H-2K* transfer; these sites contribute to the clustering of K^b-unique substitutions in comparisons involving only *H-2* sequences. This event affects the ARS.

11. $Q4^{b,k}$,$Q5^k$,$Q10^{b,k}$ 427, 430 (*H2/Qa**, *Qa/Tla**)

These *Qa* alleles resemble the *H-2* consensus, while one *H-2* allele, D^b, resembles *Q1*, *Q2*, *Q7* and *Q8*. Although an exchange event appears to have occurred here, it is not entirely clear which sequences were involved. The SCP includes no amino acid changes.

12. D^p,$K1^k$ 435, 436 (*H2/K1**)

$D^{b,k,p}$, $L^{d,q}$	ctgctc
$K^{b,q}$	.....
$Q1^k$, $Q4^{b,k}$, $Q10^{b,k}$	.....
$K1^k$	....g.
Human and bovine consensus	....g.
$Q2^k$	.....a
T13	a....g
$D^{d,q}$	gc...g (SCP #9)
$K^{k,s}$	gc...g (SCP #9)
$Q5^k$	gc...g (SCP #9)
K^d	gca.ag (SCP #9)
$Q7^{b,d}$, $Q8^b$	gca.ag (SCP #9)
Human *HLA-B51*	gc....

FIGURE 5.—Positions 314–319, illustrating SCP #9.

This cluster is an example of a case in which our modification of STEPHENS' algorithm produces poor results. D^p is clustered with $K1^k$ because they share a substitution at 436 and because each has a unique substitution at 435, but this is not solid evidence for conversion.

13. D^k,$K^{b,k,q,s}$ 504, 507 (*H2**)

Allele D^k shows a sequence typical of *H-2K*, suggesting conversion by that locus; however, the region has a pronounced "patchwork" pattern suggesting that multiple overlapping conversions have occurred here. If a K^s-like sequence acted as donor, the region of similarity is 101 bp. The event affects the ARS.

14. K^d,L^d,$Q2^k$,$Q5^k$,$Q10^{b,k}$ 536, 538, 539 (*Qa***)

This clustering is significant in comparisons including only *H-2* sequences, suggesting some transfer among *Q2*, *Q5* and *Q10* as well as a *Qa/H-2* event. The complex pattern of similarities in this part of the ARS (illustrated in Figure 6) obscures assignment of donor and recipient sequences, and also prevents detection of this SCP in comparisons involving *H-2* as well as *Qa* sequences.

15. $Q2^k$,$Q4^{b,k}$,$Q7^{b,d}$,$Q8^b$ 569, 573, 579 (*Qa***)

This clustering suggests a transfer into $Q2^k$ from an allele of the *Q4/Q7/Q8* branch. It spans a maximum of 61 bp if *Q4* was the donor, and affects the ARS.

16a. $D^{b,d,q}$,$L^{d,q}$,$K^{b,q,s}$,*T13* 593, 594 (All*)
16b. $D^{b,d,q}$,$L^{d,q}$,$K^{b,q,s}$ 590, 593, 594 (*H2/Qa****, *H2/K1****, *H2****)

Two *H-2K* alleles and two *H-2D* alleles follow the *Qa* consensus here, while the remaining *H-2* alleles have a sequence similar to *T13*. It is difficult to determine the direction or number of transfers in this

$D^{d,k}$	agag	
K^k	....	
$Q8^b$	....	
K^s	...t	
K^b	...c	
$Q7^{b,d}$	.a..	
$Q1^k$	.a.c	
$Q4^{b,k}$	.a.a	
K^q	..gc	
D^p	.cgc	
$D^{b,q,Lq}$	catt	(SCP #14)
L^d	tatt	(SCP #14)
K^d	tatt	(SCP #14)
$Q2^k,Q5^k,Q10^{b,k}$	tatt	(SCP #14)

FIGURE 6.—Positions 536–539, illustrating SCP #14.

region, which is part of the non-ARS framework of the molecule.

17a. $Q7^{b,d},Q8^b$ 32, 655, 659, 661, 666, 721 (All***)
17b. 32, 527, 528, 655, 659, 661, 666, 721 (*H2/Qa**)

A tight cluster of unique substitutions in the closely related *Q7/Q8* loci suggests a conversion among these sequences, or from an unidentified donor into their common ancestor. If the bases at positions 527 and 528 are part of this conversion, it involves a region homologous to the ARS in $\alpha 2$ and spans intron 2; otherwise it affects only $\alpha 3$.

Alpha-3 (Exon 4: 621–905):

18. $K^{b,k,q,s}$ 650, 663, 839, 863 (*H2/Qa**)

Across this region the most divergent *H-2K* allele, K^d, is quite different from all other *K* alleles; at some sites it follows the *H-2* and *Qa* consensus while the other *K* alleles differ, and at others it has unique or nearly unique substitutions. No known sequence can be identified as a donor, nor is it clear whether the exchange was into K^d or into an ancestor of the other *H-2K* alleles.

19. $D^d,K^{b,k,q,s}$ 651, 656 (*H2**)

Allele D^d shows a sequence typical of *H-2K* suggesting conversion by a *K* allele.

20a. $D^d,Q5^k,T13$ 923, 925, 926 (All**)
20b. $D^d,Q5^k$ 413, 841, 914, 922, 923, 925, 926, 934, 951, 977 (*H2/Qa*: P_{d_0}*, P_{g_0}***)

The D^d allele shows extensive homology with $Q5^k$ between positions 841 and 977, including both the unique substitutions listed here and other substitutions shared by $Q5^k$ and other *Qa* alleles or *T13*. This is a large event, apparently spanning intron 4 (not analyzed).

21. $D^{d,p}K1^k,Q5^k$ 857, 859 (All*, *H2/Qa**)

A pair of substitutions is shared by two related *H-2D* sequences, the pseudogene $K1^k$, and $Q5^k$. The surrounding highly conserved sequences give no hints to the direction of the transfer or transfers, and do not allow a meaningful estimate of event length.

trans-Membrane (Exon 5: 906–1043):

22. $D^d,K^{d,k,q,s}$ 933, 937, 941 (*H2**)

This region shows similarities between D^d, a group of related *H-2K* sequences, and $Q5^k$. The $D^d/Q5^k$ similarity may be part of SCP #20, but the *H-2K* sequences also seem to have been involved in some type of transfer here.

23. $Q1^k,Q2^k,Q5^k$ 936, 939, 972 (*Qa**)

This SCP suggests a transfer among these sequences, but direction is unclear, and as several events have apparently occurred in this region, it may represent an effect of SCPs #21, #22 and #26 rather than an independent instance of exchange.

24a. $Q4^{b,k},Q8^b$ 1107, 1114 (*H2/Qa**)
24b. 941, 1037, 1107, 1114 (*Qa**)

Four substitutions and a one-bp insertion at position 968 are shared by *Q4* and *Q8* in this region, a pattern suggesting either transfer from *Q4* to *Q8* or from another *Qa* sequence into *Q7*. If the former interpretation is correct, this appears to be a long event spanning introns 5 and 6. It has a drastic effect on the protein, introducing a frameshift which generates a stop codon.

25a. All *H-2K* alleles 969, 970, 978 (All**)
25b. 969, 970, 978, 1095, 1103 (*H2/Qa***)
25c. 20, 38, 924, 969, 970, 978, 1095, 1103 (*H2/K1*: P_{g_0}**)

Sites at which all *H-2K* alleles, but not the duplicated pseudogene $K1^k$, share unique substitutions cluster in the 3′ region of the gene. This could be due to exchanges among them, or to loss of similarity among them due to exchanges elsewhere. Alternatively, the presence of these substitutions in all *K* alleles suggests the possibility that natural selection is maintaining locus-specific characters in this region.

26a. All $K,K1^k,Q1^k$ 973, 974 (All*, *H2/Qa**)
26b. All K, $K1^k$ 161, 928, 971, 973, 974 (*H2/K1*: P_{g_0}*)

Across the same region as SCP#25 there is a cluster of sites at which *H2-K*, $H\text{-}2K1^k$ and $Q1^k$ share substitutions, and these sites are significantly clustered even in comparisons without $Q1^k$. A conversion from a *K*

or *K1* sequence into $Q1^k$, apparently spanning 15 bp, is suggested, as is a possible exchange between *K* and $K1^k$.

Cytoplasmic (Exons 6–8: 1044–1157):

27a. $Q5^k$,$Q7^{b,d}$ 1069, 1091, 1105 (*H2/Qa**)
27b. 1091, 1105, 1119 (*Qa***)

Positions grouping $Q5^k$ with *Q7* cluster in the extreme 3′ region of the gene. This may be due to SCP #24, which disrupted the normal similarity of *Q7* and *Q8* in this region, and to SCP #5, which introduced novel variants into *Q5*; or it may represent an independent event, possibly a transfer from *Q5* or *Q7*. It is 3′ of translation stop in these loci.

"Gap" partitions and unclassifiable SCPs:

28. $Q10^{b,k}$ 22 sites, gap from 109–377$_*$ (All: P_{g_0}*, *Qa/Tla*:P_{g_0}*

Substitutions unique to *Q10* have apparently been eliminated from the region between positions 109 and 377 by SCPs #3 and #6, leading to this significant gap.

29a. $Q7^{b,d}$ 78, 90, 91, 883, 950 (All: P_{g_0}*, *Qa/Tla*: P_{g_0}*)
29b. 78, 90, 91, 134, 190, 883, 950, 1034 (*H2/Qa*: P_{g_0}*, *Qa*: P_{g_0}*)

The leader and the region 3′ of position 883 show many *Q7*-unique substitutions, including an insertion; elsewhere *Q8* shares most substitutions with *Q7*. Several interpretations are possible; loss of *Q7*-like sites from *Q8* by conversions (perhaps including SCP #24), or conversions of the *Q7* ancestor after it diverged from *Q8* (perhaps including SCP #27).

30a. $Q5^k$,$Q10^{b,k}$ 10, 20, 924, 928 (*Qa/Tla*: P_{g_0}*)
30b. 10, 22, 924, 928, 1153, 1154 (*Qa*: P_{g_0}*)

These loci cluster only in the leader and the extreme 3′ end of the gene suggesting either a disruption of similarity in the remainder of the gene by recombination with other sequences (for example SCP #5) or two exchanges between the *Q5* and *Q10* loci.

31a. $Q5^k$ 40 sites, gap from 284–396 (All: P_{d_0}****, P_{g_0}*)
31b. 52 sites, gap from 284–396 (*Qa/Tla*: P_{g_0}*)
31c. 55 sites, gap from 785–1153 (*H2/Qa*: P_{d_0}*, P_{g_0}***)
31d. 79 sites spanning 4–977 (*Qa****)

Two regions are lacking in $Q5^k$-unique substitutions, which are abundant elsewhere. The first gap, between positions 284 and 396, may represent the traces of SCPs #5 and #9. The second gap, in the extreme 3′ region of the gene, may reflect the numerous 3′ SCPs involving $Q5^k$: #20, #21, #23 and #27.

32a. $Q1^k$ 24 sites spanning 24–570 (All**)
32b. 34 sites, gap 643–1010 (*H2/Qa*: P_{g_0}*)
32c. 32 sites, gap 606–973 (*Qa/Tla*: P_{g_0}*)
32d. 44 sites, gap 606–973 (*Qa*: P_{g_0}*)

Sites specific to $Q1^k$ are frequent in the 5′ region of the gene but nearly absent from the 3′ region; it is unclear whether this represents introduction of 5′ divergence by conversion or loss of 3′ divergence.

33. $Q2^k$ 20 sites spanning 105–691 (*Qa/Tla**)

Sites unique to $Q2^k$ are absent in the 3′ region of the gene, apparently due to SCPs #14, #15 and #23.

34. $Q4^{b,k}$ 22 sites spanning 17–714 (*Qa**)

Sites unique to *Q4* are absent in the 3′ region of the gene, perhaps due to SCP #24.

DISCUSSION

Thirty-four significantly clustered partitions (SCPs) representing potential instances of genetic exchange were identified in analysis of 24 murine class I major histocompatibility sequences. Some of these clusters appear to be due to other causes: SCP #12 may represent an artifact of the statistical algorithm, and SCPs #28-#34, which were identified as gaps in otherwise unclustered partitions, may be "background" to other identified events rather than representing gene conversions in their own right. SCP #25 (clustering of *K*-locus specific bases) may be due to selection rather than to conversion. The following analysis will therefore concentrate on the remaining 25 SCPs, which we believe are most likely due to genetic exchange in the form of gene conversion or recombination.

Length of conversion tracts: For 16 of the 25 SCPs apparently representing conversions, it was possible to estimate the length of the putative conversion by examining the surrounding nucleotides and assuming that the conversion spanned the entire region in which the putative donor and recipient were identical. (In some of the remaining cases the putative conversion extended past one end of the gene or into an intron; in others, the conversion was apparently from an unknown sequence, providing no way to estimate length.) These estimates should be regarded as extremely tentative. Conversion length can readily be overestimated (due to misidentification of bases which are actually identical by descent as part of a conversion) or underestimated (due to subsequent point mutations or gene conversions which overlay part of a conversion tract). Since events extending into introns or off the ends of the gene could not be measured, these estimates are biased toward shorter events. The

length of the longest exon, 276 bp, represents the maximum-length event which could possibly be measured.

Of the 16 events whose lengths could be directly estimated, 14 apparently spanned less than 70 bp; one spanned 101 bp; and one spanned 214 bp. These lengths are comparable to the K^{bm} and K^{hm} spontaneous mutants, which spanned less than 200 bp (NATHENSON *et al.* 1986; VOGEL *et al.* 1988), and are also similar to the gene conversions observed when *H-2* genes were inserted into yeast chromosomes (WHEELER *et al.* 1990).

Regions of the gene involved in conversions: SCPs were detected throughout the coding sequence, affecting all exons except for the very short exon 8 (see Figure 2). Slightly more than half (14/25) of the detected SCPs were in the highly variable exons 2 and 3, which encode the antigen binding site. Furthermore, not all of the events in these exons affected the antigen binding site: one SCP (#11) was completely silent and three (#6, #7, #16) affected framework portions of the molecule rather than the surfaces thought to be involved in antigen presentation and T cell recognition. This suggests that the mechanism of conversion in these genes is not limited to the antigen binding site, but produces conversions across the length of the gene. Lack of intron sequence for some alleles and difficulties in alignment prevented us from testing for conversion events in the introns, but inspection of the sequences clearly indicates that conversion-like patterns exist in intron as well as exon sequences (data not shown).

Several SCPs (#31–#34) involved the absence of sites specific to a particular *Qa* locus in the 3′ end of the gene (beyond the stop codon in most *Qa* alleles). The lack of such clusters in *H-2*, and their abundance in *Qa*, suggests that some mechanism other than the gene conversion observed across the gene may account for them. One possibility is that selection is acting differently on the 5′ and 3′ ends of *Qa* genes. A priori, we would have expected the *Qa* loci, not all of which express the most 3′ exons, to be enriched for locus-specific substitutions in these exons (which are highly conserved in functional *H-2* loci). The actual finding–absence of 3′ locus-specific substitutions–is surprising, and points to a possible high rate of between-locus conversion in the 3′ region of *Qa* loci, tending to homogenize the sequences. Alternatively, there may be positive selection for locus-specific substitutions in the 5′ region of these genes, perhaps due to some functional specialization.

Directionality of conversions: Although STEPHENS' algorithm (STEPHENS 1985) simply identifies clustering of similarities without giving a direction of exchange, examination of the sequences and their similarity trees (Figure 3) can often be used to deduce direction. For example, a sequence shared by several *Qa* loci and by just one of the *H-2D* alleles was probably transferred from *Qa* to *H-2* rather than the reverse; this deduction is strengthened if the region of similarity includes substitutions which are typical of *Qa* rather than of *H-2*.

Thirteen clusters allow deduction of donor and recipient sequences. Of these, five involve only *Qa* sequences and three only *H-2* sequences. Three apparently involve transfer from *Qa* into *H-2*; one from *H-2* into *Qa*; and one from *Tla* into *Qa*. These findings do not support the contention that information flows only from nonclassical to classical sequences (BREGEGERE 1983), especially as several of the clusters for which donor and recipient cannot be unambiguously determined appear to have involved *H-2* to *Qa* transfers.

Two additional clusters appear to represent transfers into *H-2*, and one into *Qa*, from sequences outside the current dataset. As more sequences become available, donors for these events may be found; however, it is likely that some of the donor sequences are no longer present in modern populations.

CONCLUSIONS

Twenty-five of the clusters detected using STEPHENS' algorithm (STEPHENS 1985) appear to be *bona fide* gene conversion events. They are similar in both size and position to the $H\text{-}2K^{bm}$ and $H\text{-}2K^{hm2}$ spontaneous mutations (NATHENSON *et al.* 1986; GELIEBTER and NATHENSON 1988; VOGEL *et al.* 1988) (see Figure 2) and to conversion events detected when the mouse genes are carried in yeast chromosomes (WHEELER *et al.* 1990). However, unlike the spontaneous mutations detected on the basis of allograft rejection, the events are distributed throughout the length of the class I gene, affecting the leader sequence, conserved framework regions within the antigen binding site, and the 3′ structural exons as well as the hypervariable antigen presentation site positions. Thus, the gene conversion mechanism does not appear to exhibit a preference for functional regions of the molecule. In addition, the events involved not only *H-2*, but also *Qa* and *Tla* loci, confirming the occurrence of gene conversion in the evolution of nonclassical as well as classical genes. We detected one SCP in which transfer was apparently from *H-2* to *Qa* and several more in which such transfer was probable. Since the 24 sequences compared contain only a proportion of the genetic information available among class I alleles, and the ability to detect events is diminished by overlapping recombination or mutation, the gene conversion events identified here are likely to represent only a fraction of the sequence exchanges which have shaped the evolutionary history of the class I gene family.

Three other possible explanations of clustering

should be considered. First, if STEPHENS' algorithm is used to compare two groups of sequences with very different patterns of variability, it may produce spurious evidence for genetic exchange. However, this effect should produce locus-specific or region-specific clusters, since patterns of variability are presumably characteristic of the locus or region involved. Only one SCP (#25: clustering of substitutions specific to *H-2K* in the transmembrane region) which could be explained in this way was observed. Furthermore, a large proportion of the SCPs were significantly clustered in comparisons involving only *H-2* or only *Qa* sequences, which should avoid this effect. Finally, a cross-species comparison indicated only one SCP suggesting conversion among distantly related species (human and mouse; SCP #9) while if variation in degree of relatedness were a major source of falsely significant SCPs, this comparison should have generated many such SCPs.

A second source of clustering could be strong selection for a particular variant, causing it to evolve independently in several lineages. Although this mechanism could account for a proportion of our SCPs (particularly SCP #9, in which the same polymorphism is observed in species as distant from mouse as cow and human), it cannot explain the shared silent substitutions that characterize many of them. Convergent evolution would not be expected to produce the same silent mutations in each lineage.

A third possible explanation for clustering is the persistence, perhaps due to natural selection, of ancestral similarities in a region where other alleles have diverged. Although this mechanism could account for a few of the clusters observed, it will not explain SCPs such as #20, where a unique sequence is found in just one *H-2* and one *Qa* allele. If the unique sequence is in fact ancestral, gene conversion must still be invoked to explain why all other sequences share the same, non-ancestral form.

Identification of evidence for gene conversion is not necessarily inconsistent with proposals that class I polymorphism is very old (FIGUEROA, GUNTHER and KLEIN 1988; LAWLOR *et al.* 1988; MAYER *et al.* 1988). Balancing selection could result in the persistence of polymorphic epitopes for extremely long periods; gene conversion, rather than eliminating such epitopes, may actually contribute to their persistence by reassorting them with newly arisen, advantageous variants. This is consistent with the observation that several SCPs cannot be explained by single instances of gene conversion because they occur in too many groups of unrelated alleles (for example SCP #6, where D^p, $K1^k$, and several *Qa* alleles share clustered substitutions), although it is also possible that certain regions of the gene are preferentially converted. The gene conversions observed are short enough, compared to the length of the gene, that they could occur fairly frequently without disrupting the overall similarity of related alleles, consistent with the observation that certain human and chimpanzee class I alleles have remained very similar since the two lineages diverged (LAWLOR *et al.* 1988).

Gene conversion has been invoked as a diversifying mechanism among *H-2* loci. If gene conversion is occurring among *Qa* and *Tla* loci, how is the relative monomorphism of these loci maintained? Several observations argue that *Qa* and *Tla* genes are not under the same selective constraints as the *H-2* genes: (i) *Qa* and *Tla* loci are relatively nonpolymorphic (CHEN *et al.* 1987; WEISS 1987; WATTS *et al.* 1989), (ii) although some at least are expressed, *Qa* and *Tla* gene products do not appear to function as classical T cell restriction elements (CHEN *et al.* 1987; ROBINSON 1987), and (iii) significantly greater numbers of unique nucleotide substitutions are found in *Qa* compared to *H-2* sequences (HUGHES and NEI 1988). This suggests that the force which maintains *H-2* polymorphism and *Qa* monomorphism in the face of gene conversion may be natural selection. Just as with point mutations, gene conversions which introduce variability into the *H-2* genes may be selectively advantageous and preserved by natural selection, while similar conversions into *Qa* or *Tla* loci may be neutral or detrimental and thus less likely to be propagated.

Gene conversion is usually thought of as a homogenizing mechanism in multigene families. However, it can also generate new alleles if the length of the conversions is less than the length of the gene, as it apparently is in *H-2*; in this case, conversions can produce new alleles by reassorting variability present in existing alleles. In the *H-2* antigen presenting site, selection for variability may favor conversions which produce novel combinations of epitopes. Elsewhere in the *H-2* gene and in the *Qa* and *Tla* genes, the absence of this selection, combined in those loci which produce functional products with purifying selection to maintain function, may favor conversions which reduce variability and preserve locus-specific residues. In this connection, it is interesting to note that the only locus-specific clustering in the *H-2* genes (SCP #25) is in the conserved 3′ region of the gene.

The variety of evolutionary forces acting on the class I genes make them a useful model for many processes occurring elsewhere in the genome. As in *Qa* and *Tla*, gene conversion could have an homogenizing effect in gene systems such as hemoglobin (COLLINS and WEISSMAN 1984), where apparently minor perturbations of protein structure can result in drastic differences in function. On the other hand, gene conversion could contribute to the generation of diversity in gene systems for which allelic variability is advantageous, such as the class I, class II (HOOD,

STEINMETZ and MALISSEN 1983; KLITZ, THOMSON and BAUR 1986), immunoglobulin (REYNAUD *et al.* 1987), and T-cell receptor gene families. Considerable work remains to be done in describing both the physical mechanisms and evolutionary consequences of gene conversion in complex multigene families.

We thank WARREN GISH and CHRIS MEACHAM for computing assistance, LI JIN for providing the phylogeny software, and PETER PARHAM, MINNIE MCMILLAN, CLAY STEPHENS, and CHRIS WHEELER for comments on the manuscript. This work was supported by National Institute of Health grants CA37099 (R.S.G.), GM35326 (G.T.) and Pregraduate Training Grant GM07232-10 (S.W.).

LITERATURE CITED

BJORKMAN, P. J., M. A. SAPER, B. SAMRAOUI, W. S. BENNETT, J. L. STROMINGER and D. C. WILEY, 1987a Structure of the human class I histocompatibility antigen, *HLA-A2.* Nature **329:** 506–512.

BJORKMAN, P. J., M. A. SAPER, B. SAMRAOUI, W. S. BENNETT, J. L. STROMINGER and D. C. WILEY, 1987b The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. Nature **329:** 512–518.

BREGEGERE, F., 1983 A directional process of gene conversion is expected to yield dynamic polymorphism associated with stability of alternative alleles in class I histocompatibility antigens gene family. Biochimie **65:** 229–237.

CHEN, Y.-T., Y. OBATA, E. STOCKERT, T. TAKAHASHI and L. J. OLD, 1987 *Tla*-region genes and their products. Immunol. Res. **6:** 30–45.

COLLINS, F. S., and S. M. WEISSMAN, 1984 The molecular genetics of human hemoglobin. Proc. Nucleic Acid. Res. Mol. Biol. **31:** 315–462.

ENNIS, P. D,. A. P. JACKSON and P. PARHAM, 1990 Rapid cloning of *HLA-A,B* cDNA using the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **87:** 2833–2837.

FIGUEROA, F., E. GUNTHER and J. KLEIN. 1988 MHC polymorphism pre-dating speciation. Nature **335:** 265–267.

FISHER, D. A., S. W. HUNT III and L. HOOD, 1985 Structure of a gene encoding a murine thymus leukemia antigen, and organization of *Tla* genes in the BALB/c mouse. J. Exp. Med. **162:** 528–545.

FLAGHERTY, L., E. ELLIOT, J. TINE, A. WALSH and J. WATERS, 1990 The genes and products of the *Qa* subregion of the murine major histocompatibility complex. CRC Crit. Rev. Immunol. (in press).

GELIEBTER, J., and S. G. NATHENSON, 1988 Microrecombinations generate sequence diversity in the murine major histocompatibility complex: Analysis of the K^{bm3}, K^{bm4}, K^{bm10} and K^{bm11} mutants. Mol. Cell. Biol. **8:** 4342–4352.

GUSSOW, D., R. S. REIN, I. MEIHER, W. DEHOOG, G. H. A. SEEMAN, F. M. HOCKSTENBACK and H. L. PLOEGH, 1987 Isolation, expression, and the primary structure of *HLA-Cw1* and *HLA-Cw2* genes—evolutionary aspects. Immunogenetics **25:** 313–322.

HOLMES, N., P. ENNIS, A. M. WAN, D. W. DENNEY and P. PARHAM, 1987 Multiple genetic mechanisms have contributed to the generation of the *HLA-A2/A28* family of class I MHC molecules. J. Immunol. **139:** 936–941.

HOOD, L., M. STEINMETZ and B. MALISSEN, 1983 Genes of the major histocompatibility complex of the mouse. Ann. Rev. Immunol. **1:** 529–568.

HUGHES, A. L., AND M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

HUGHES, A. L., and M. NEI, 1989 Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol. Biol. Evol. (in press).

KLEIN, J., 1986 *Natural History of the Major Histocompatibility Complex.* John Wiley & Sons, New York.

KLITZ, W., G. THOMSON and M. P. BAUR, 1986 Contrasting evolutionary histories among tightly linked HLA loci. Am. J. Hum. Genet. **39:** 340–349.

KUHNER, M. K., and R. GOODENOW, 1989 DNA sequences of mouse *H-2* and *Qa* genes. Immunogenetics **30:** 458–464.

LAWLOR, D. A., F. E. WARD, P. D. ENNIS, A. P. JACKSON and P. PARHAM, 1988 *HLA-A* and *B* polymorphism predates the divergence of humans and chimpanzees. Nature **335:** 268–271.

MAYER, W. E., M. JONKER, D. KLEIN, P. IVANYI, G. VAN SEVENTER and J. KLEIN, 1988 Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. EMBO J. **7:** 2765–2774.

MORITA, T., C. DELARBRE, M. KRESS, P. KOURILSKY and G. GACHELIN, 1985 An *H-2K* gene of the t^{w32} mutant at the *T/t* complex is a close parent of an $H\text{-}2K^q$ gene. Immunogenetics **21:** 367–383.

MORSE, R., S. WATTS, C. GELBER and R. S. GOODENOW, 1990 The *Qa* region genes and antigens of the murine major histocompatibility complex, in *Immunology and Biotechnology in Biomedical Research, Series in Current Concepts in Immunogenetics*, Vol. 2, edited by P. TYLE *et al.* Verlag Chemie International Publications, Berlin (in press).

NATHENSON, S. G., J. GELIBTER, G. M. PFAFFENBACH and R. A. ZEFF, 1986 Murine major histocompatibility complex class I mutants: molecular analysis and structure-function implications. Annu. Rev. Immunol. **4:** 471–502.

PARHAM, P., C. E. LOMEN, D. A. LAWLOR, J. P. WAYS, N. HOLMES, H. L. COPPIN, R. D. SALTER, A. M. WAN and P. D. ENNIS, 1988 The nature of polymorphism in *HLA-A*, *-B*, and *-C* molecules. Proc. Natl. Acad. Sci. USA **85:** 4005–4009.

PARHAM, P. D., A. LAWLOR, C. E. LOMEN and P. D. ENNIS, 1989 Diversity and diversification of *HLA-A,B,C* alleles. J. Immunol. **142:** 3937–3950.

PEASE, L. R., 1985 Diversity in *H-2* genes encoding antigen-presenting molecules is generated by interactions between members of the major histocompatibility complex gene family. Transplantation **39:** 227–231.

REYNAUD, C. A., V. ANQUEZ, H. GRIMAL and J. C. WEILL, 1987 A hyperconversion mechanism generates the chicken light chain preimmune repertoire. Cell **48:** 379–388.

ROBINSON, P. J., 1987 Structure and expression of polypeptides encoded in the mouse *Qa* region. Immunol. Res. **6:** 46–56.

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

SODOYER, R., M. DAMOTTE, T. L. DELOVITSH, J. TRUCY, B. R. JORDAN and T. STRACHAN, 1984 Complete nucleotide sequence of a gene encoding a functional human class I histocompatibility antigen (*HLA-Cw3*). EMBO J **3:** 879–885.

STEPHENS, J. C. 1985 Statistical methods of DNA sequence analysis: Detection of intragenic recombination or gene conversion. Mol. Biol. Evol. **2:** 539–556.

VOGEL, J. M., A. C. DAVIS, D. M. MCKINNEY, W. J. MARTIN and R. S. GOODENOW, 1988 Molecular characterization of the C3HfB/HeN $H\text{-}2K^{tm2}$ mutation: Implications for the molecular basis of alloreactivity. J. Exp. Med. **168:** 1781–1800.

WATTS, S., A. C. DAVIS, B. GAUT, C. WHEELER, L. HILL and R. S. GOODENOW, 1989 Organization and structure of the *Qa* genes of the major histocompatibility complex of the C3H mouse: implications for *Qa* function and class I evolution. EMBO J. **8:** 1749–1759.

WEISS, E. H., 1987 Molecular biology of the mouse *Q* region. Immunol. Res. **6:** 179–191.

WHEELER, C. J., D. MALONEY, S. WATTS, J. VOGEL, J. TOWNER, D. BALDWIN, J. RUFER, M. ARBEITMAN, H. CHAN, S. FOGEL and R. S. GOODENOW, 1990 Microconversion between murine *H-2* genes integrated into yeast, in *Transgenic Mice and Mutants in MHC Research*, edited by I. K. EGOROV and C. S. DAVID. Springer-Verlag, Berlin.

Communicating editor: A. G. CLARK