Ancient Interlocus Exon Exchange in the History of the *HLA-A* Locus

Austin L. Hughes and Masatoshi Nei

Center for Demographic and Population Genetics, The University of Texas Health Science Center, Houston, Texas 77225

Manuscript received November 22, 1988
Accepted for publication March 31, 1989

ABSTRACT

The major histocompatibility complex (MHC) in humans and chimpanzees includes three classical class I loci, *A*, *B* and *C*, which encode glycoproteins expressed on the surface of all nucleated cells. There are also several nonclassical class I loci including *E*, which have more limited expression. By analyzing published sequences, we have shown that in exons 4 and 5, *A* locus alleles from both humans and chimpanzees are much more similar to *E* than to *B* or *C* alleles, whereas in exons 2 and 3 alleles from all three classical class I loci are much more similar to each other than any one is to *E*. We propose that some 20 million years ago, interlocus recombination led to the formation of a hybrid gene in which exons 2 and 3 were derived from the original *A* locus and exons 4 and 5 were derived from the *E* locus. The fact that such an ancient event can still be detected suggests that interlocus recombination is rare in the MHC and does not significantly contribute to MHC polymorphism, which is known to be extremely high. The present finding, however, supports Gilbert's idea that exons in a gene may occasionally be replaced by those from another gene in the evolutionary process.

IN a recent study (HUGHES and NEI 1988), we compared rates of synonymous and nonsynonymous (amino acid altering) nucleotide substitution in various regions of mouse and human class I major histocompatibility complex (MHC) genes. We found that the rate of nonsynonymous substitution substantially exceeds that of synonymous substitution in the 57 codons encoding the antigen recognition site (BJORKMAN *et al.* 1987a, b) of the class I molecule, whereas in other regions of the same genes the synonymous rate generally exceeds the nonsynonymous rate. We interpreted these results as evidence of overdominant selection acting on the antigen recognition site (HUGHES and NEI 1988). However, our study revealed an anomalous feature that could not be explained by the overdominance hypothesis alone. That is, in exon 4, which encodes the $\alpha 3$ domain of the class I molecule, the *HLA-A* gene in humans was found to be highly divergent from the *HLA-B* and *-C* genes.

In order to explain this observation, we proposed that in the past interlocus exon exchange had occurred. We hypothesized that the ancestral *A* locus received exon 4 from a locus not closely related to it or to *B* and *C*. At that time we had no idea about the donor locus. Here we show that the divergent nature of exon 4 of *HLA-A* can be explained as a result of past recombination with an *HLA-E*-like gene.

DNA SEQUENCES USED

The human classical class I loci, *HLA-A*, *-B*, and *-C*, encode MHC molecules, which are expressed on all nucleated cells. Their main function is to present intracellularly processed foreign peptides to cytotoxic T cells (KLEIN 1986). The chimpanzee genome also includes three classical class I loci, designated as *ChLA-A*, *-B*, and *-C* (MAYER *et al.* 1988). Typically in mammals, there are several nonclassical class I genes in addition to the classical genes; nonclassical genes have a more limited tissue expression and their function is not known (HOWARD 1987). It has been suggested that most or all nonclassical genes have no function and, even though they are sometimes expressed, they are effectively pseudogenes (HOWARD 1987; KLEIN and FIGUEROA 1986). A few nonclassical human genes have been sequenced. One such locus has been designated *HLA-E* (KOLLER *et al.* 1988).

Classical and nonclassical class I genes have a similar structural organization. Exon 1 encodes the leader peptide, and exons 2, 3 and 4 encode the three extracellular domains, $\alpha 1$, $\alpha 2$ and $\alpha 3$, respectively. The 57 residues of the antigen recognition site are located in the $\alpha 1$ and $\alpha 2$ domains. Exon 5 encodes the transmembrane portion of the molecule, and exons 6, 7 and 8 encode the cytoplasmic tail. The number of amino acid residues in the cytoplasmic tail varies considerably among different loci. In this paper we compared published nucleotide sequences for exons 2–5 of eight human and two chimpanzee class I alleles. Exons 1 and 6–8 were excluded from our analysis because alignment among distantly related genes was uncertain in these regions. Intron sequences were not used since several published sequences were for cDNA. The sequences analyzed included two *HLA*-A alleles (KOLLER and ORR 1985; N'GUYEN *et al.* 1985), two *HLA-B* alleles (KOTTMANN *et al.* 1986;

```
                                                                                    20                                                                              40
       Exon 2           +       +       +                                                   +       +       +
A3     GGC TCC CAC TCC ATG AGG TAT TTC TTC ACA TCC GTG TCC CGG CCC GGC CGC GGG GAG CCC CGC TTC ATC GCC GTG GGC TAC GTG GAC GAC ACG CAG TTC GTG CGG TTC GAC AGC GAC GCC
Ch-A   --- --- --- --- --- --- --- --- --- -A- --C --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
B13    --- --- --- --- --- --- --- --- --- -A- --C G-- A-- --- --- --- --- --- --- --- --- --- --- A-- --- --- --- --- --- --- --C --- --- --- A-- --- --- --- --- ---
Ch-B   --- --- --- --- --- --- --- --- --- -A- --C G-- --- --- --- --- --- --- --- --- --- --- --- T-A --- --- --- --- --- --- --- --- --- --- T-- --- --- --- --- ---
Cw1    T-- --- --- --- --- -A- --- --- --- --- --- --- --- --- --T --- --- --A --- --- --- --- --- T-A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
E      --- --- --- --- T-- -A- --- --- CA- --T --- --- --- --- --- --- --- --- --- --- --- --- --- T-T --- --- --- --- --- --- --C --- --- --- --C --- --- -A- --- ---

                                                                                    60                                                                              80
                                                                        +   +   +       +   +   +   +   +   +   +   +   +   +   +   +   +   +   +   +   +           +
A3     GCG AGC CAG AGG ATG GAG CCG CGG GCG CCG TGG ATA GAG CAG GAG GGG CCG GAG TAT TGG GAC CAG GAG ACA CGG AAT GTG AAG GCC CAG TCA CAG ACT GAC CGA GTG GAC CTG GGG ACC
Ch-A   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- A-- --- --- TCC G-- --- --- --- --- --- --- --- --- ---
B13    A-- --T -C- --- --- -C- --C --- --- --A --- --- --- --- --- --- --- --- --- --- --- -G- --- --- -A- -TC TCC --- A-- A-C A-- --- --- T-- --- -A- A-- --- C-C ---
Ch-B   --- --T -C- --A GA- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- -GA --- --- -A- -TC TCC --- A-- A-C G-- --- --- T-- --- -A- AG- --- C-- -A-
Cw1    --- --T -C- --A GG- --- --- --- --- --- --- G-G --- --- --- --- --- --- --- --- --- -G- --- --- -A- --G TAC --- CG- --- G-- --- --- --- --- --- AG- --- C-- -A-
E      --- --T -C- --- --- -T- --- --- --- --- --- --G --- --- --- --- T-A --- --- --- --- -G- --- --- --- -GC -CC -G- -A- ACC G-- --- -T- TT- --- --- A-- --- C-- --G

                                                                                   100                                                                             120
        +   +       +                          Exon 3           +       +       +                                                           +       +
A3     CTG CGC GGC TAC TAC AAC CAG AGC GAG GCC|GGT TCT CAC ACC ATC CAG ATA ATG TAT GGC TGC GAC GTG GGG TCG GAC GGG CGC TTC CTC CGC GGG TAC CGG CAG GAC GCC TAC GAC GGC
Ch-A   --- --- --- --- --- --- --- --- --- -A-|--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
B13    GC- -T- C-- --- --- --- --- --- --- ---|--G --- --- --T TGG --- -CG --- --- --- --- --- C-- --- C-- --- --- --- C-- --- --- --- C-T AAC --- TTA --- --- --- ---
Ch-B   --- --- --- --- --- --- --- --- --- ---|--- --- --- -T- --- --- -GG --- --C --- --- --- A-- --- C-- --- --- --- C-- --- --- --- --T GA- --- T-- --- --- --- ---
Cw1    --- --- --- --- --- --- --- --- --- ---|--G --- --- --- C-- --- TGG --- -G- --- --- --G C-- --- C-C --- --- --- C-- --- --- --- --T GAC --- T-- --- --- --- ---
E      --- --- --- --- --- --T --- --- --- ---|--G --- --- --- C-G --- TGG --- C-- --- --- --- C-- --- C-C --- A-- --- --- --- --- --- --T GAA --- TT- --- --- --- ---

                                                                                   140                                                                             160
                                                                                                +       +   +   +       +   +   +   +       +   +   +   +   +   +
A3     AAG GAT TAC ATC GCC CTG AAC GAG GAC CTG CGC TCT TGG ACC GCG GCG GAC ATG GCG GCT CAG ATC ACC AAG CGC AAG TGG GAG GCG GCC CAT GAG GCG GAG CAG TTG AGA GCC TAC CTG
Ch-A   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -C- --- --A --- --- --- --- --- --- --- --- --- --- --- --- -C- --- --- --- C-- --- --- --- ---
B13    --- --- --- --- --- --- --- --- --- --- A-- --C --- --- --- --- --- -CC --- --- --- --- --- C-- -T- --- --- --- --- --- -G- -T- --- --- --- C-- --- --- --- ---
Ch-B   --- --- --- --- --- --- --- --- --- --- A-- --C --- --- --- --- --- -CC --- --- --- --- --- C-- --- --- --- --- --- --- -G- TG- --- --- --- C-- --- --- --- ---
Cw1    --- --- --- --- --- --- --- --- --- --- --- --C --- --- --C --- --- -CC --- --- --- --- --- C-- --- --- --- --- --- --- -G- --- --- --- --- CG- --- --- --- ---
E      --- --- --T C-- A-- --- --T --- --- --- --- --C --- --- --- -T- --- -C- --- --- --- --- T-- G-- -AA --- -CA A-T -AT --- TC- --- --- --- --C CA- --- --- --- ---

                                                                                   180                                                                             200
        +   +   +       +   +   +       +       +                                              Exon 4
A3     GAT GGC ACG TGC GTG GAG TGG CTC CGC AGA TAC CTG GAG AAC GGG AAG GAG ACG CTG CAG CGC ACG|GAC CCC CCC AAG ACA CAT ATG ACC CAC CAC CCC ATC TCT GAC CAT GAG GCC ACC
Ch-A   --G --- CG- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---|--- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
B13    --G --- GA- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- G--|--- --- --A --- --- --C G-- --- --- --- --- --- --- --- --- --- --- ---
Ch-B   --G --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- GT-|--- --- --A --- --- --- G-- --- --- --- --- --- --- --- --- --- --- ---
Cw1    --G --- --- --- --- --C --- --- --- --- --- --- --- --- --- --- --- -GC --- --- --- G--|--A -A- --A --- --- --C G-- --- --- --T --- G-- --- --- --- --- --- ---
E      --A -A- --A --- --- --- --- --- -A- -A- --- --- --- --G --- --- --- --- --- -TT -A- CT-|--G --- --A --- --- --C G-- --T --- --- --- --- --- --- --- --- --- ---

                                                                                   220                                                                             240
A3     CTG AGG TGC TGG GCC CTG GGC TTC TAC CCT GCG GAG ATC ACA CTG ACC TGG CAG CGG GAT GGG GAG GAC CAG ACC CAG GAC ACG GAG CTC GTG GAG ACC AGG CCT GCA GGG GAT GGA ACC
Ch-A   --- --- --- --- --- --- --- --- --- --- --- -T- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --- --- ---
B13    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --A --T --- --- --T --- --T --- --- --- --A --A --- --A --- A-- ---
Ch-B   --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --C --- --- --A --T --- --- --C --- --T --- --- --- --A --A --- --A --- A-- ---
Cw1    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- T-- --- --- --- --- --A --T --- --- --C --- --T --- --- --- --- --A --- --A --- --- ---
E      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -A- --- --- --- -G- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---

                                                                                   260                                                                             280
                                                                                                                                           Exon 5
A3     TTC CAG AAG TGG GCG GCT GTG GTG GTG CCT TCT GGA GAG GAG CAG AGA TAC ACC TGC CAT GTG CAG CAT GAG GGT CTG CCC AAG CCC CTC ACC CTG AGA|TGG GAG CTG TCT TCC CAG CCC
Ch-A   --- --- --- --- --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---|--- --- -C- --- --- --- ---
B13    --- --- --- --- --A --- --- --- --- --- --- --- --A --- --- --- --- --A --- --- --A --- --- --- --G --- --G --- --- --- --- --- ---|--- --- -CA --- --- --- T--
Ch-B   --- --- --- --- --A --- --- --- --- --- --- --- --A --- --- --- --- --A --- --- --A --- --- --- --G --- --G --- --- --- --- --- ---|--- --- -CA --- --- --- T--
Cw1    --- --- --- --- --A --- --- A-- --- --- --- --- --A --- --- --- --- --G --- --- --- --- --C --- --G --- --G G-- --- --- --- --- ---|--- --- -C- --- --- --- ---
E      --- --- --- --- --A --- --- --- --- --- --- --- --- --- --- --- --- --G --- --- --- --- --- --- --G --A --- G-- --- G-- --- --- ---|--- A-- -C- G-- --- --- ---

                                                                                   300
A3     ACC ATC CCC ATC GTG GGC ATC ATT GCT GGC CTG GTT CTC CTT GGA GCT GTG ATC ACT GGA GCT GTG GTC GCT GCC GTG ATG TGG AGG AGG AAG AGC TCA
Ch-A   --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
B13    --- G-- --- --- --- --- --T G-- --- --- --- -C- G-- --A -C- -T- --- G-- -TC --- --- --- --- --- --T --- --- --T --- --- --- --- ---
Ch-B   --- --- --- --- --- --- --T G-- --- --- --- -C- G-- --A -C- -T- --- G-- -TC --- --- --- --- --- --T --- --- --T --- --- --- --- ---
Cw1    --- --- --- --- --- --- --- G-- --- --- --- -C- G-- --G -CT -TC C-A GCT GTC --- --- --- --G --- -TT --- --- --T --- --- -G- --- ---
E      --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- T-- --- G-- T-- --- --- --- --T --- --T --- --A --- --- -A- --- --- ---
```

FIGURE 1.—Six DNA sequences of exons 2–5 from human and chimpanzee class I MHC loci. *A3*, *B13* and *Cw1* are human classical genes. *Ch-A* and *Ch-B* stand for the chimpanzee classical genes *CH-A108* and *CH-B2*, respectively, whereas *E* is a human nonclassical gene. Codons encoding residues in the antigen recognition site (ARS) are marked above with "+." "-" indicates identity with the *A3* sequence.

ZEMMOUR *et al.* 1988), two *HLA-C* alleles (GUSSOW *et al.* 1987), one *ChLA-A* allele, one *ChLA-B* allele (MAYER *et al.* 1988), and *HLA-E* gene.

RESULTS AND DISCUSSION

Figure 1 shows DNA sequences for exons 2–5 from the *HLA-E* and from selected human and chimpanzee classical class I genes. Inspection of these sequences suggests that the relationship between the *E* gene and the classical genes for exons 2 and 3 is different from that for exons 4 and 5. In exons 2 and 3, one can see many unique substitutions which set *E* apart from the classical sequences. In fact, *E* shows unique amino-acid residues at 31 positions; these residues are not shared by any of the classical genes illustrated in Figure 1 or by any of the other classical genes in our sample. Of these 31 residues unique to *E*, 18 are in the antigen recognition site of the MHC molecule, which includes many highly polymorphic residues. Exon 4 shows fewer nonsynonymous differences among sequences, as is expected in a highly conserved domain. However, there are many synonymous substitutions that are shared by most *A* and *E* sequences but not by *B* and *C*. For example, at codons 225, 237, 253, and 267, the third-position nucleotide is identical for *E* and all *A* sequences (including those not illustrated in Figure 1), whereas all *B* and *C* sequences

TABLE 1

Nucleotide substitutions per site (*d*) (above diagonal) and synonymous substitutions per synonymous site (d_S) (below diagonal) in exons 2 and 3 in comparisons among human and chimpanzee class I alleles

	A3	*Aw24*	*Ch-A108*	*B13*	*B44*	*Ch-B2*	*Cw1*	*Cw2*	*E*
A3		.059	.030	.128	.121	.104	.108	.121	.179
Aw24	.015		.075	.115	.119	.119	.119	.141	.186
Ch-A108	.039	.035		.128	.119	.106	.128	.131	.186
B13	.129	.101	.125		.055	.077	.121	.117	.175
B44	.118	.108	.117	.052		.063	.119	.115	.191
Ch-B2	.104	.091	.104	.071	.066		.079	.087	.168
Cw1	.109	.110	.134	.119	.091	.079		.053	.163
Cw2	.150	.155	.163	.155	.121	.088	.062		.189
E	.208	.161	.233	.168	.184	.188	.177	.217	

The prefix "*Ch-*" denotes chimpanzee alleles; others are human genes.

(including those not illustrated) share a different third-position nucleotide. Nonsynonymous differences are more numerous in exon 5 than in exon 4. In these regions *A* and *E* share many second and third position substitutions. At amino acid positions 288, 292, and 293, all *A* and *E* sequences share an amino acid residue differing from that of all *B* and *C* (including those not illustrated).

In order to quantify the pattern of nucleotide substitution among the genes in our sample, we estimated the number of nucleotide substitutions per site (*d*) and the number of synonymous substitutions per synonymous site (d_S; NEI 1987) in pairwise comparisons among the 10 genes. d_S's were estimated by method I of NEI and GOJOBORI (1986). We computed these values separately for exons 2–3 and exons 4–5. Table 1 shows the results for exons 2–3. In exons 2–3, *E* is quite divergent from classical genes of the human and chimpanzee in terms of both total substitutions (*d*) and synonymous substitutions (d_S). In the case of exons 4–5, however, *A* locus alleles from both the human and chimpanzee are in general more similar to *E* than they are to *B* or *C* locus alleles (Table 2). The divergence of *A* and *E* from *B* and *C* is particularly marked in the case of synonymous substitutions. The amino acid sequence is conserved at most positions in the transmembrane region (encoded by exon 5) and is highly conserved in the $\alpha3$ domain (encoded by exon 4). Thus the rate of synonymous substitution exceeds the nonsynonymous rate in exons 4 and 5 (HUGHES and NEI 1988, and unpublished data). For this reason, synonymous substitutions may give a clearer picture of evolutionary relationships among the genes than total substitutions.

Using the distance values shown in Tables 1 and 2, we constructed evolutionary trees using the unweighted pair group method (UPGMA; SNEATH and SOKAL 1973), to examine the relationships among the genes in our sample. We constructed trees separately for exons 2–3 and exons 4–5. Figure 2, A and B, shows the trees based on total nucleotide differences. In the case of exons 2–3 (Figure 2A), the *B* and *C* loci appear to be more closely related to each other than either is to the *A* loci of either the human or chimpanzee, but *E* forms a separate cluster, apart from all the classical loci. By the method of NEI, STEPHENS and SAITOU (1985), the branch point connecting *E* with the cluster of classical loci is significantly different from that connecting the *A* cluster to the *B–C* cluster ($t = 6.31$; $P < 0.001$). The UPGMA tree for exons 4–5 presents a very different topology (Figure 2B). Here, although *B* and *C* alleles again cluster together, *A* alleles cluster with *E* rather than with *B* and *C* alleles. In this case, the branch point connecting the *B–C* cluster to the *A–E* cluster is significantly different from that connecting *A* and *E* ($t = 4.42$; $P < 0.001$). Note that, while each of the *A*, *B* and *C* loci forms a separate cluster in Figure 2B as in Figure 2A, the details of topology within these clusters differ between the two figures. However, these differences are not statistically significant because the distances between the branch points within clusters are very small.

Figure 2, C and D, shows the trees based on synonymous substitutions only. In the case of exons 2–3, *E* again forms a separate cluster apart from the classical loci (Figure 2C). Here also, the branch point between the *E* cluster and the classical gene cluster is significantly different from that between the two major clusters of classical genes ($t = 3.29$; $P < 0.01$). Note that *B* alleles are grouped with *A* alleles rather than with *C* alleles in this case. However, the difference between the two branch points involved is not statistically significant. In the case of exons 4–5 (Figure 2D), *E* again clusters with *A* alleles, the difference between the branch point connecting *A* and *E* and that connecting the *A–E* cluster to the *B–C* cluster again being significant ($t = 5.19$; $P < 0.001$).

In order to confirm the evolutionary relationships revealed by these UPGMA trees, we constructed trees using the neighbor-joining method (SAITOU and NEI

TABLE 2

Nucleotide substitutions per site (*d*) (above diagonal) and synonymous substitutions per synonymous site (d_S) (below diagonal) in exons 4 and 5 in comparisons among human and chimpanzee class I alleles

	A3	*Aw24*	*Ch-A108*	*B13*	*N44*	*Ch-B2*	*Cw1*	*Cw2*	*E*
A3		.026	.013	.092	.089	.089	.121	.121	.064
Aw24	.075		.029	.086	.089	.089	.127	.121	.075
Ch-A108	.031	.086		.098	.095	.095	.127	.121	.072
B13	.248	.234	.276		.010	.010	.078	.067	.109
B44	.221	.234	.248	.031		.010	.078	.072	.112
Ch-B2	.249	.235	.277	.031	.020		.078	.067	.112
Cw1	.280	.295	.310	.168	.156	.182		.015	.121
Cw2	.306	.291	.337	.142	.154	.155	.020		.121
E	.097	.132	.132	.206	.206	.234	.222	.247	

The prefix "*Ch-*" denotes chimpanzee alleles; others are human genes.

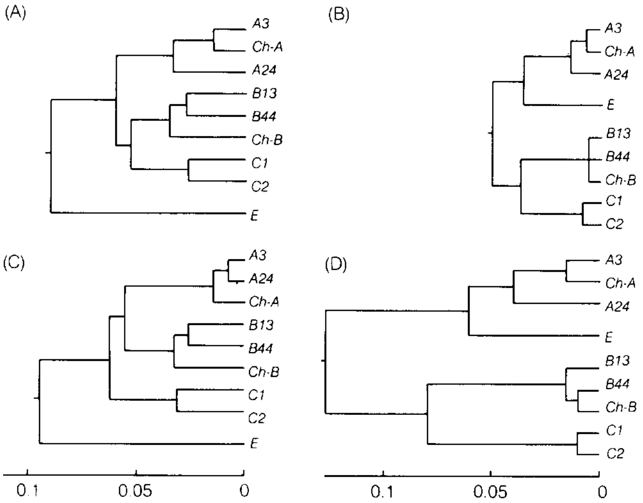

FIGURE 2.—Evolutionary trees showing relationships among exons 2–3 (A and C) and exons 4–5 (B and D) of human and chimpanzee class I MHC alleles. Trees A and B are based on total nucleotide difference (*d*), while trees C and D are based on synonymous differences only (d_S). The prefix "*Ch*⁻" indicates chimpanzee alleles. Abbreviated allele designations: *Ch-A* = *Ch-A108*, *A24* = *Aw24*, *Ch-B* = *Ch-B2*, *C1* = *Cw1*, and *C2* = *Cw2*.

1987) from the same set of data. This method is based on the principle of minimum evolution and is insensitive to differences in the rate of evolution among different lineages. The trees constructed by this method showed the same relationships among loci as the UPGMA trees (the trees not shown).

The extremely long third intron (between exons 3 and 4) of class I MHC genes contains numerous repeats (RONNE *et al.* 1985) and has been proposed as a recombinational hotspot in the mouse (WATTS *et al.* 1987). Although other introns could be aligned, KOLLER and ORR (1985) considered alignment of this intron between *A* and *C* alleles impossible, suggesting past recombination in this intron. We hypothesize that interlocus recombination occurred in this intron, joining exons 2 and 3 of an ancestral *A* gene with exons 4 and 5 of an *E*-like gene (Figure 3). All known *A* locus alleles in both humans and chimpanzees are apparently descended from this hybrid allele. Because the hybrid sequence is observed in both human and chimpanzee *A* alleles, the *A*–*E* recombinational event apparently occurred prior to separation of the human and chimpanzee lineages (5–7 million years ago). However, it must have occurred after the separation of *HLA-A* from the *HLA-B* and *-C* loci (about 40 million years ago; KLEIN and FIGUEROA 1986), since the alleles from the latter loci do not share the hybrid sequence. This suggests that the recombination event occurred some time between 5 and 40 million years ago. If we assume that the *E* gene in our sample is descended from a gene closely related to the one involved in recombination with *A* and that synonymous substitution occurs at a rate proportional to time, we can estimate the age of the recombination event by the extent of synonymous substitution between exons 4–5 of *A* and *E* in comparison with that between *A* and *B* plus *C* (Figure 3). This method yields an estimate of about 20 million years for the age of the recombination event.

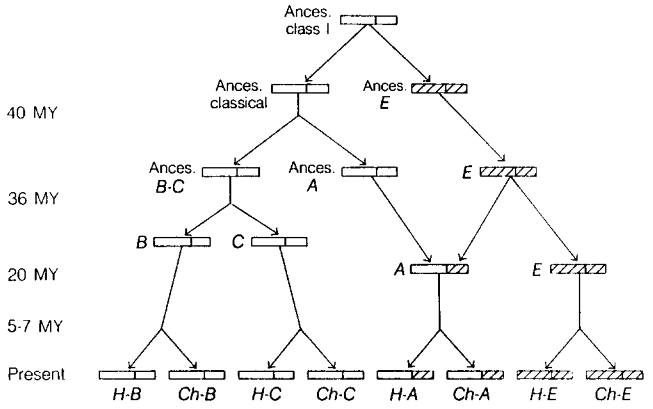

FIGURE 3.—Schematic representation of a hypothetical recombinational event involving the ancestral *A* locus and an *E*-like locus. The vertical line in each gene represents the position of the third intron, proposed site of recombination. The figure is not intended to represent relative chromosome positions of the ancestral *A* and *E* genes.

As an alternative hypothesis to explain the data presented here, one might argue that the divergence of exons 2–3 of *E* from those of classical class I genes is due to selection on the recognition site of exons 2 and 3. On this hypothesis, *A* and *E* would actually be closely related, but their close relationship would be obscured by selection for amino acid substitutions in the recognition site. This hypothesis, however, does not explain why the exons 2–3 of *E* are markedly different from those of classical genes at synonymous sites (Figure 2C), which are presumably not affected by selection. Furthermore, in a tree based on exons 2 and 3 excluding the 57 recognition site codons, the *E* again clusters apart from the classical class I sequences (the tree not shown).

When introns were first discovered, it was suggested that new genes might evolve by exon shuffling (GILBERT 1978). By now, a number of examples of this process have been reported (DOOLITTLE 1985; SÜDHOFF *et al.* 1985a, b), but most of them are based on homologies between exons of very distantly related genes. The event of exon shuffling reported here differs in that the genes involved are closely related and that the donor gene can be identified. In the present case, however, it is not clear whether any change in function accompanied the *A–E* recombination. A case of exon shuffling was also reported by HOLMES and PARHAM (1985), but this represents an event of intragenic recombination.

The present data support the occurrence of an ancient exon exchange in the history of human and chimpanzee *A* alleles, but it should be noted that interlocus exon exchange is an extremely rare event. This is obvious because we could identify the trace of recombination only once in the entire history of the *HLA-A*, *B* and *C* loci that spans tens of millions of years. A number of authors (*e.g.*, WEISS et al. 1983; MELLOR 1987) proposed that interlocus recombination or gene conversion is an important mechanism of maintenance of MHC polymorphism. However, the exon exchange detected here has little to do with MHC polymorphism, because a large portion of allelic variation in MHC loci is caused by variation in the antigen recognition site of exons 2 and 3. Furthermore, recombination or gene conversion cannot explain the high rate of nonsynonymous nucleotide substitution in comparison with the synonymous rate. As suggested by HUGHES and NEI (1988, 1989), the extremely high level of polymorphism at MHC loci (80–90% heterozygosity) seems to be mainly due to overdominant selection.

This study was supported by grants from the National Institutes of Health and the National Science Foundation.

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.

DOOLITTE, R. F., 1985 The genealogy of some recently evolved vertebrate proteins. Trends Biochem. Sci. **10:** 233–237.

GILBERT, W., 1978 Why genes in pieces? Nature **271:** 501.

GUSSOW, P., R. S. REIN, I. MEIJER, W. DE HOOG, G. H. SEEMAN, F. M. HOCHSTENBACH 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., and P. Parham, 1985 Exon shuffling in vivo can generate novel HLA class I molecules. EMBO J. **4:** 2849–2854.

HOWARD, J. C., 1987 MHC organization of the rat: evolutionary considerations, pp. 397–411. in *Evolution and Vertebrate Immunity*, edited by G. KELSOE and D. H. SCHULZE, University of Texas Press, Austin.

HUGHES, A. L., and M. 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 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. Proc. Natl. Acad. Sci. USA **86:** 958–962.

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

KLEIN, J., and F. FIGUEROA, 1986 Evolution of the major histocompatibility complex. CRC Crit. Rev. Immunol. **6:** 295–389.

KOLLER, B. H., and H. T. ORR, 1985 Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. J. Immunol. **134:** 2727–2733.

KOLLER, B. H., D. E. GERAGHTY, Y. SHIMIZU, R. DEMARS and H. T. ORR, 1988 *HLA-E*: a novel HLA class I gene expressed in resting T lymphocytes. J. Immunol. **141:** 897–904.

KOTTMANN, A. H., G. H. A. SEEMAN, H. D. GUESSOW and M. H. ROOS, 1986 DNA sequence of the coding region of the *HLA-B44* gene. Immunogenetics **23:** 396–400.

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.

MELLOR, A., 1987 Molecular genetics of class I genes in the mammalian histocompatibility complex. Oxf. Surv. Eukaryotic Genes **3:** 95–140.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol. Biol. Evol. **3:** 418–426.

NEI, M., J. C. STEPHENS and N. SAITOU, 1985 Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. **2:** 66–85.

N'GUYEN, C., R. SODOYER, J. TRUCY, T. STRACHAN and B. R. JORDAN, 1985 The *HLA-Aw24* gene: sequence, surroundings, and comparison with the *HLA-A2* and *HLA-A3* genes. Immunogenetics **21:** 479–489.

BONNE, H., E. WIDMARK, L. RASK and P. A. PETERSON, 1985 Intron sequences reveal evolutionary relationships among major histocompatibility complex genes. Proc. Natl. Acad. Sci. USA **82:** 5860–5864.

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

SNEATH, P. H. A., and R. R. SOKAL, 1973 *Numerical Taxonomy.* W. H. FREEMAN, San Francisco.

SÜDHOF, T. C., J. L. GOLDSTEIN, M. S. BROWN and D. W. RUSSELL, 1985a The LDL receptor gene: a mosaic of exons shared with different proteins. Science **228:** 815–822.

SÜDHOF, T. C., D. W. RUSSELL, J. L. GOLDSTEIN and M. C. BROWN, 1985b Cassette of eight exons shared by genes for LDL receptor and EGF precursor. Science **228:** 893–895.

WATTS, S., J. M. VOGEL, W. D. HARRIMAN, T. ITOH, H. J. STRAUSS and R. S. GOODENOW, 1987 DNA sequence analysis of the C3H H-2K^k and H-2D^k loci: evolutionary relationships to H-2 genes from other mouse strains. J. Immunol. **139:** 3878–3885.

WEISS, E. H., A. L. MELLOR, L. GOLDEN, K. FAHRNER, E. SIMPSON, J. HURST and R. A. FLAVELL, 1983 The structure of a mutant H-2 gene suggests that the generation of polymorphism in H-2 genes may occur by gene conversion-like events. Nature **301:** 671–674.

ZEMMOUR, J., P. P. ENNIS, P. PARHAM and B. DUPONT, 1988 Comparison of the structure of *HLA-Bw47* to *HLA-Bw13* and its relationship to 21-hydroxylase deficiency. Immunogenetics **27:** 281–287.

Communicating editor: J. AVISE