0022-1767/92/14812-4043$02.00/0
THE JOURNAL OF IMMUNOLOGY

Vol. 148, 4043–4053, No. 12, June 15, 1992
Printed in U.S.A.

HLA-J, A SECOND INACTIVATED CLASS I HLA GENE RELATED TO HLA-G AND HLA-A

Implications for the Evolution of the HLA-A-Related Genes[1,2]

GERALD MESSER,* JACQUELINE ZEMMOUR,† HARRY T. ORR,‡ PETER PARHAM,[3†] ELISABETH H. WEISS,* AND JOHN GIRDLESTONE[§]

*From the *Institut für Immunologie, Universität München, W-8000 München 2, F.R.G.; †Departments of Cell Biology, and Microbiology and Immunology, Stanford University, Stanford, CA 94305-5400; ‡Institute of Human Genetics, University of Minnesota, Box 198, UMHC, Minneapolis, MN 55455; and the §MRC Laboratory of Molecular Biology, Cambridge CB2 2QH, U.K.*

Ragoussis and co-workers (*Genomics 4:301*) previously described a class I HLA gene (now designated HLA-J) that maps to within 50 kb of HLA-A. The nucleotide sequences of three HLA-J alleles are reported here. Comparison of the nucleotide sequences of HLA-J alleles shows this gene is more related to HLA-G, A, and H than to HLA-B, C, E, and F. All four alleles of HLA-J are pseudogenes because of deleterious mutations that produce translation termination either in exon 2 or exon 4. Apart from these mutations, the predicted proteins have structures similar to those of HLA-A, B, and C molecules. There is, however, little polymorphism at HLA-J and none at functional positions of the Ag-recognition site. The polymorphism is less than found for HLA-H another HLA-A-related pseudogene. HLA-J appears, like HLA-H, to be an inactivated gene that result from duplication of an Ag-presenting locus related to HLA-A. Nucleotide sequence comparisons show that the HLA-A, H, J, and G genes form a well defined group of "HLA-A-related" loci. Evolutionary relationships as assessed by construction of trees suggest the four modern loci: HLA-A, G, H, and J were formed by successive duplications from a common ancestral gene. In this scheme one intermediate locus gave rise to HLA-A and H, the other to HLA-G and J.

The HLA class I region extends over 5 megabases of the short arm of chromosome 6 and contains at least 17 homologous class I genes, pseudogenes, and gene fragments (1–3). Among these, the classical class I genes—HLA-A, B, and C—show widespread tissue expression, a highly developed polymorphism and function in the presentation of peptide Ag to $CD8^+$ T lymphocytes. Like HLA-A, B, and C, the products of three other genes—HLA-E, F, and G—associate with β_2-m, but in contrast these molecules show restricted tissue distribution, low diversity, and are of unknown function (4–6). In addition to class I genes that encode β_2-m-associated H chains, there are a number of pseudogenes. Characterization of six alleles of the class I pseudogene HLA-H revealed both low diversity and a structure, apart from two deleterious mutations, that is remarkably similar to that of Ag-presenting HLA genes: in particular HLA-A (7, 8). This suggested HLA-H represents a once functional Ag-presenting gene that had subsequently been inactivated.

Ragoussis et al. (9) recently mapped a "novel" class I HLA gene (cda12) to be within 50 kb of the HLA-A locus. Here we report sequences of three alleles of this locus and an analysis of their polymorphism. This locus, which has been designated HLA-J (10), is related to HLA-A but less so than HLA-H. As for HLA-H, all alleles of HLA-J are pseudogenes caused by a single base pair of deletions in exons 1 and 4 encoding the leader sequence and the α_3 domain, and exhibit little diversity or polymorphism. HLA-J shows a close relationship with HLA-G. Sequence comparisons suggest a scheme for the evolution of the HLA-A-related genes.

MATERIALS AND METHODS

Isolation and sequencing of HLA-J clones. HLA-J alleles were isolated from three cell lines: cd (HLA-A2; B27, B51; Cw2, Cw3) clone cda12; Molt4 (HLA-A1, 25; B57, B18; Cw2) clone Molt4px; and LCL 721 (HLA-A1,2; B8, 5) clone 59Kbd. A Molt4 λ2001 genomic library, kindly provided by N. Migone, was screened at low stringency with nick-translated insert of the HLA-B7 cDNA, pDP001 ((American Type Culture Collection, Bethesda, MD) designation), kindly provided by S. Weissman. Molt4px was one of several clones that did not anneal at high stringency with probes representing the 3′ untranslated region of HLA-A or B alleles from Molt4 (11). The λ-DNA was sonicated and cloned into Mp18 and the resulting library was screened at low stringency against the insert from a pUC19 clone covering the HLA-A1 gene from Molt4 (12). Positive clones were sequenced by the dideoxy method until both strands had been read. A cosmid library was made from the PBL of individual cd. Isolation of the cda12 clone was as described (9) and sequenced as shown in Figure 1a. Isolation and sequencing of 59Kbd was as described for the HLA-F gene (13).

mRNA analysis. S1 nuclease protection experiments were performed as described earlier (11), using 15-μg aliquots of total RNA from the Molt4 derivative YHHH, which had been cultured with or without 1000 U/ml rIFN-α (a gift of Hoffman la Roche, Nutley, NJ) for 20 h. Single strand probes for HLA-A and B were derived from

Received for publication October 8, 1991.
Accepted for publication March 24, 1992.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
[1] This work was supported by Grants SFB217 from the DFG (G. M. and E. W.) and AI24258 and HL47153 from the National Institutes of Health, Bethesda, MD (J. Z. and P. P.)
[2] The sequence data presented in this article have been submitted to the EMBL/GenBank Data Libraries under the accession numbers M80468 (Molt4px), M80469 (cda12), and M80470 (59Kbd).
[3] Address correspondence and reprint requests to Dr. Peter Parham, Department of Cell Biology, Sherman Fairchild Building, Stanford University, Stanford, CA 94305-5400.

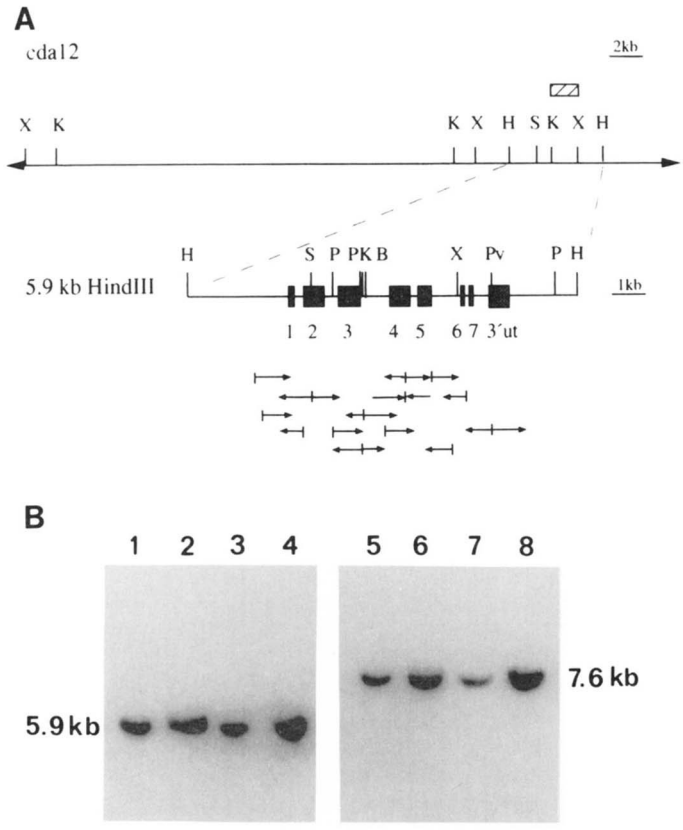

Figure 1. *A*, restriction map of an HLA-J-containing cosmid, cda12, and the 5.9-kb *Hin*dIII subclone obtained from it. *Closed boxes* indicate exons. Restriction sites shown are *H*, *Hin*dIII; *K*, *Kpn*I; *S*, *Sal*I; *X*, *Xho*I; *B*, *Bgl*II; *P*, *Pst*I, and *Pv*, *Pvu*II. The 1.5-kb *Kpn*I/*Xho*I fragment used as the *cda12*-specific probe is indicated. The sequencing strategy of subclones in M13 using fragments obtained with the restriction enzymes: *Eco*RI, *Sac*I, and *Sma*I is shown at the *bottom*. *B*, Southern blot analysis of HLA-J. Genomic DNA was digested with *Hin*dIII (*lanes 1* to *4*) or *Pvu*II (*lanes 5* to *8*), and 10 μg (*lanes 1*, *3*, *5*, and *7*) and 15 μg (*lanes 2* and *6*) and 18 μg (*lanes 4* and *8*) were separated on a 0.7% agarose gel, transferred to Hybond N⁺ (Amersham & Buchler), and hybridized with the 1.5-kb *Kpn*I/*Xho*I fragment of cda12. DNA was obtained from cells with HLA-types: LG2, A2: B27: Cw1 (*lanes 1* and *5*); CD, A2: B27,51: Cw2,3 (*lanes 2* and *6*); TY, A11:B35: Cw4; (*lanes 3* and *7*); WW, A2,3: B27,44: Cw2,5 (*lanes 4* and *8*).

M13 cDNA clones M4-201 and M4-117 (11), and for HLA-J from an M13 shotgun clone generated for the sequencing of Molt4px.

Evolutionary analysis. Trees for the classical and nonclassical class I HLA loci were constructed using the UPGMA[4] method (14). Comparison of the whole gene was based on the 5′ flanking region (~200 bp upstream from the ATG codon), all exons and introns, and 100 bp of the 3′ UTR. The 5′ part of the gene covers the 5′ flanking region through exon 3, whereas the 3′ part spans intron 3 through the 3′ UTR.

RESULTS

HLA-J is most closely related to HLA-A and HLA-G. Screening cosmid and genomic libraries derived from three human cell lines, with either an HLA-A-specific probe or a general class I HLA probe, resulted in the isolation of novel and related class I genes. The cda12, Molt4px, and 59Kbd clones derive from the cd, Molt4, and LCL 721 cell lines, respectively. These class I genes are all contained on *Hin*dIII fragments of 5.8 to 5.9 kb as detected by Southern blot (Fig. 1b), and determination and comparison of their nucleotide sequences revealed these genes to be highly homologous to each other. Moreover, these genes are also related to the recently published partial sequence (exons 2 and 3) of a gene called DAN2 (15). In fact the DAN2 sequence is identical to exons 2 and 3 of both the cda12 and 59Kbd genes.

The overall structure of HLA-J alleles is similar to that of other class I HLA genes: the exon-intron organization is the same and many of the nucleotide positions conserved in other class I genes are found in HLA-J (Fig. 2). The upstream control regions include normal class I gene enhancer, IFN response, and promoter elements. Furthermore, HLA-J genes of the Molt4 cell line were shown to be transcribed and inducible with IFN-α (Fig. 3, *A* and *B*). In contrast, when the cda12 gene was transfected into mouse L929 or P815 cells no HLA-J mRNA was detected.

Pairwise comparisons of the coding region sequences show differences in 1 to 8 nucleotides between cda12, Molt4px, and 59Kbd whereas comparison with the alleles of other class I HLA loci gives differences in excess of 100 nucleotides (Table I). These properties clearly indicate that these genes (and DAN2) are alleles of a class I HLA locus distinct from HLA-A, B, C, E, F, G, and H. Following the convention of alphabetic order for the designation of HLA genes, but to avoid confusion between I and 1 this locus has been termed HLA-J (10). From previous studies with the cda12 clone, HLA-J has been mapped to within 50 kb of HLA-A (9).

Comparison with the classical genes, HLA-A, B, and C, shows HLA-J is most closely related to HLA-A (Table I), consistent with its map position and its detection by HLA-A-specific probes. It also exhibits a comparable level of similarity with HLA-H, which is also related and mapped close to HLA-A (7, 25). Comparison with the nonclassical class I HLA genes, HLA-E, F, and G, showed HLA-G was the most closely related of all class I HLA genes to HLA-J, whereas HLA-E and F were the most divergent. HLA-G differs from HLA-J by 112 substitutions compared with 118–133 for HLA-A alleles (Table I). The similarities with HLA-G include a cluster of 6 substitutions in a stretch of 28 bp in the carboxyl-terminal half of α_2 (codons 147–156), which are uniquely shared by HLA-J and HLA-G. Thus HLA-J has similarities with particular classical, nonclassical and nonfunctional class I HLA loci: HLA-A, HLA-G, and HLA-H.

The similarities of HLA-J with HLA-A, G, and H can be appreciated further from examination of the patterns of nucleotide substitution at those "locus-specific" positions that permit distinction of HLA-A, B, and C alleles. At these positions all alleles of a single HLA-A, B, or C locus have an identical nucleotide but there are differences between the loci (26). For almost all of these positions, two of the loci have the identical nucleotide and the third is divergent. From a total of 58 such positions, HLA-J alleles are identical to HLA-A at 40 positions, to HLA-B at 24 positions, and to HLA-C at 20 positions (Table II). On this basis the relatedness of HLA-J to HLA-A is comparable with that seen with HLA-H. Further comparisons with HLA-E, F, and G show they are all more closely related to HLA-A than to either HLA-B or C and that HLA-J shares the greatest number of "locus-specific positions"—45—with HLA-G.

The majority of locus-specific positions derive from the 3′ half of the class I gene. From previous comparisons it

[4] Abbreviations used in this paper: UPGMA, unweighted pair group method using arithmetic averages.

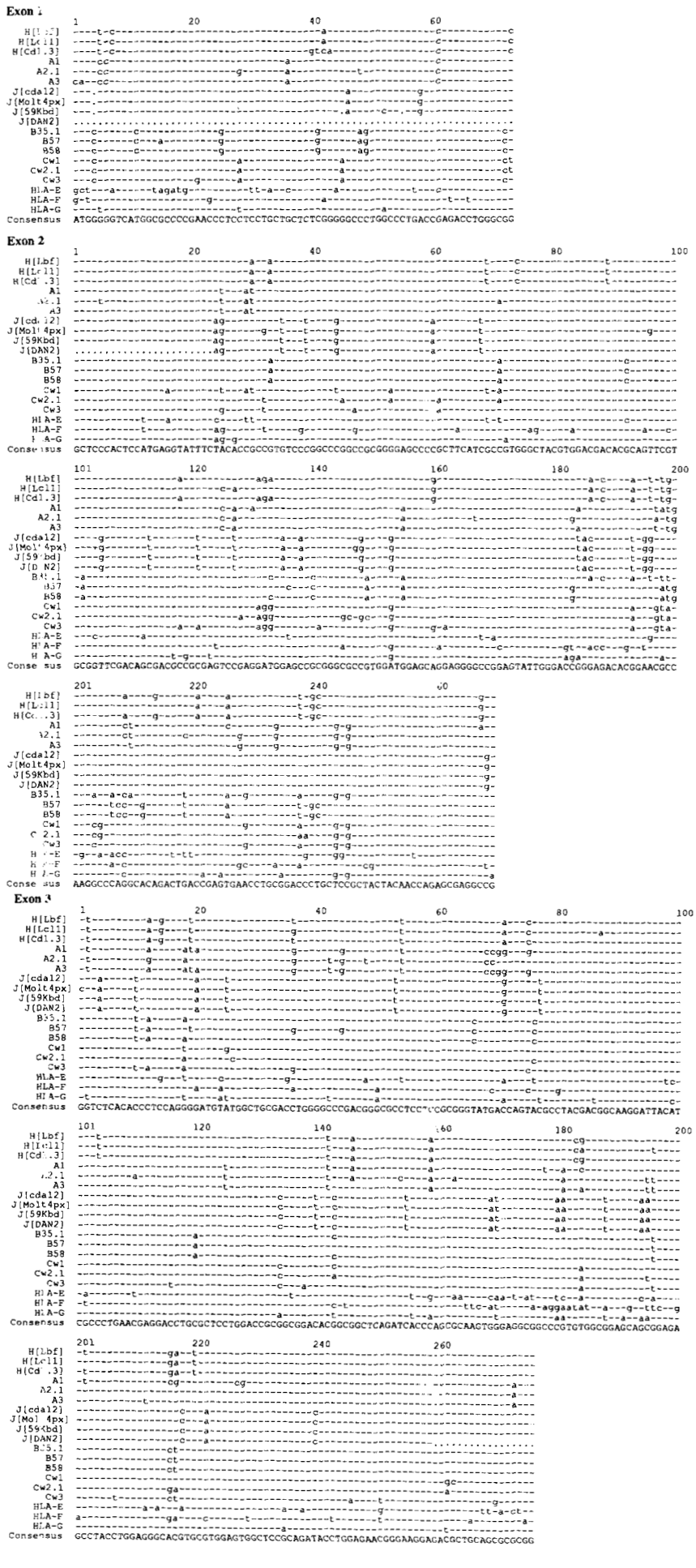

Figure 2. Comparison of the nucleotide sequences of the coding regions of four HLA-J alleles with HLA-A, B, C, E, F, G, and H alleles. The gaps introduced for optimal homology are indicated by dots. The sequences are from Zemmour et al. H(Lbf), H(Lc11)) (7); Girdlestone (A1) (12); Koller and Orr (A2.1) (16); Strachan et al. (A3) (17); Ooba et al. (B35.1) (18); Isamat et al. (B57) (19); Ways et al. (B58) (20); Güssow et al. (Cw1, Cw2) (21); Sodoyer et al. (Cw3) (22); Koller et al. (HLA-E) (23); Geraghty et al. (HLA-F) (13); Geraghty et al. (HLA-G) (24); H(Cd1.3) is an HLA-H allele isolated from the cd cosmid library. HLA-A, B, and C alleles chosen for comparison are those for which complete gene sequences—exons and introns—are available.

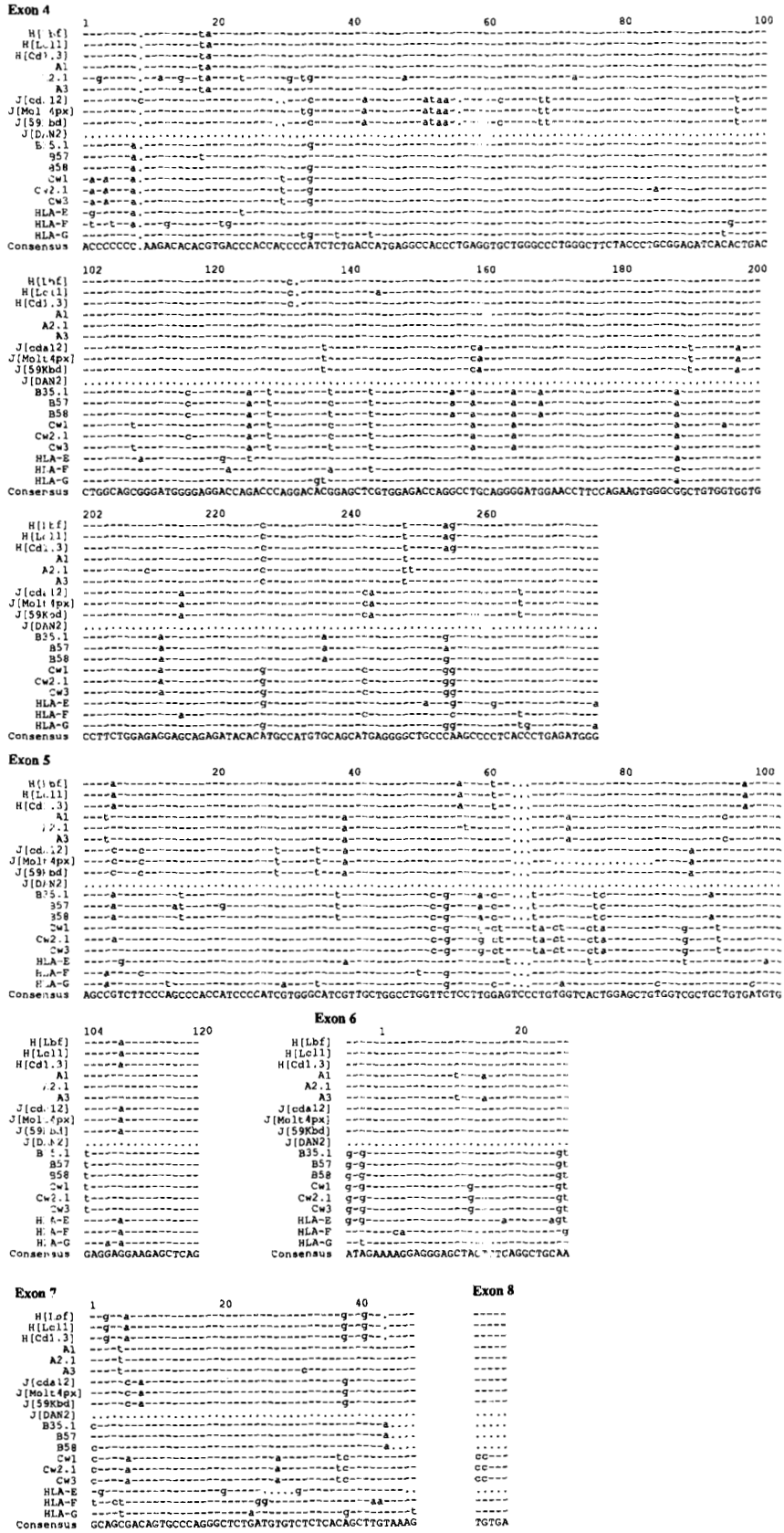

Figure 2. continued.

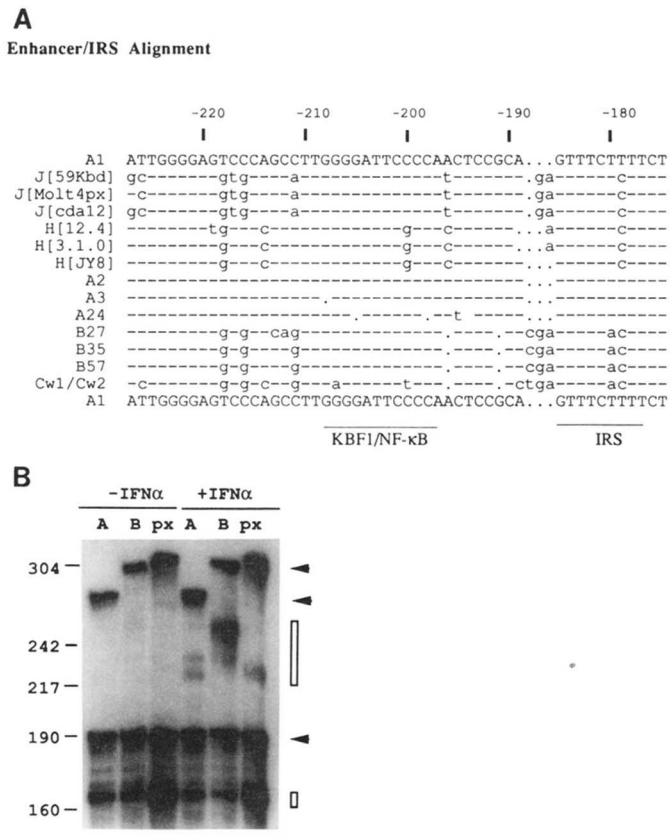

Figure 3. *A*, comparison of the enhancer A and IRS sequences of the 5′ promoter region of HLA-J to HLA-A, B, C, and H alleles. Sequences are compared with those of HLA-A1 that are shown at the *top* and *bottom*. *Numbers* indicate the position of the sequence relative to the initiation codon ATG. *B*, induction of HLA-J mRNA by IFN-α. Nuclease protection analysis was performed on RNA from the Molt4 derivative YHHH (11). All three probes were designed to complement portions of the 3′ untranslated regions of their respective transcripts because these sequences exhibit maximal divergence between HLA-J and other known HLA class I genes. *Arrowheads* indicate full length undigested cDNA, *open bars* indicate the protected species, and the *numbers* represent length in base pairs. The *lower arrowhead* and *bar* represent the probe for adenosine deaminase, included as a control for RNA levels.

became clear there are qualitative differences in the pattern of sequence diversity in the 5′ (exons 1 to 3) and 3′ (exons 4 to 8) parts of the gene (26). This pattern is also seen for HLA-J: in the 5′ half of the entire gene there is comparable sequence divergence between HLA-J and other class I HLA genes with a slightly closer relationship to HLA-A and somewhat greater divergence from HLA-E and F (Table IIIA). In the 3′ part of the gene, however, HLA-J is clearly more related to HLA-A, G, and H and divergent from HLA-B, C, E, and F (Table IIIB). The similarity between HLA-J and HLA-G is greater than that between HLA-J and either HLA-A or HLA-H.

Evolution of HLA-A related alleles. To assess further the relationships of HLA-J to other class I HLA loci we constructed evolutionary trees employing the UPGMA method (Fig. 5). Trees obtained with the complete sequence—exons, introns, and untranslated regions—showed that HLA-A, G, H, and J form a group of HLA-A-related loci that are readily distinguishable from HLA-B, C, E, and F (Fig. 5*A*). Furthermore, within this group of four loci HLA-A is closer to HLA-H, whereas HLA-J is closer to HLA-G.

When just the 3′ part of the gene is submitted to this analysis (Fig. 5*B*), a tree with similar topology to that observed for the complete gene (Fig. 5*A*) is seen. From previous comparisons it appeared that the 3′ part of class I HLA genes is under different selection than the 5′ part and was more likely to reflect the ancestry of the genes (26). These results suggest that HLA-J and G were formed by duplication of a common ancestor, as were HLA-A and H.

Trees made from the 5′ part of the gene (Fig. 5*C*) have a different topology from those seen with the 3′ part or the whole gene. In these trees HLA-A, B, C, and H form a group of four more closely related loci. This grouping is probably a reflection of selection for diversity in Ag presentation and support the hypothesis that HLA-H is a more "recently" inactivated Ag-presenting locus than HLA-J.

HLA-J is a Relatively Nonpolymorphic Pseudogene. Although showing a high degree of nucleotide sequence

TABLE I

Pairwise nucleotide comparisons of coding region of class I sequences (from exon 1 through exon 8)

	HLA-H[Lc11]	HLA-A2	HLA-J			HLA-B58	HLA-Cw1
			cda12	Molt4px	59Kbd		
HLA-H(Lbf)	14[a]	103	127	131	128	111	119
HLA-H(Lc11)	—	95	124	128	125	109	121
HLA-H(Cd1.3)	18	106	130	134	131	113	120
HLA-A1	80	50	123	126	124	113	124
HLA-A2	95	—	132	132	133	116	127
HLA-A3	80	42	118	121	119	115	120
HLA-J(cda12)	124	132	—	7	1	138	135
HLA-J(Molt4px)	128	132	7	—	8	138	135
HLA-J(59Kbd)	125	133	1	8	—	138	136
HLA-B35	108	113	138	138	138	18	83
HLA-B57	104	116	143	144	143	16	101
HLA-B58	109	116	138	138	138	—	91
HLA-Cw1	121	127	135	135	136	91	—
HLA-Cw2	124	136	141	140	142	88	31
HLA-Cw3	121	136	143	142	144	91	39
HLA-E	149	148	170	173	171	155	152
HLA-F	147	156	149	152	150	158	164
HLA-G	118	113	112	112	112	126	123

[a] Values are the number of nucleotide differences obtained by pairwise comparison of the designated sequences. The gaps introduced for optimal homology of sequences were considered as identities.

TABLE II

Locus-specific nucleotide positions

Exon	Position	Codon	Locus-Specific Nucleotides									
			A	H	B	C	J			E	F	G
							Molt4px	cda12	59Kbd			
1	4	−23	G	G	C	C	.	.	.	G	G	G
	6	−23	C	G	G	G	G	G	G	G	G	G
	36	−13	A	G	G	G	G	G	G	C	G	G
	61	−4	C	C	G	G	G	G	G	C	G	G
2	102	35	C	C	A	C	C	C	C	C	C	C
	125	42	C	T	T	T	T	T	T	T	T	T
	153	52	A	A	A	G	G	G	G	A	G	G
	155	52	A	G	A	G	G	G	G	G	G	G
3	2	91	T	T	G	G	G	G	G	G	G	T
	125	132	T	C	C	C	C	C	C	C	C	C
	142	138	T	T	C	C	C	C	C	C	C	C
4	2	183	C	C	C	A	C	C	C	G	T	C
	4	184	C	C	C	A	C	C	C	C	C	C
	8	185	C	C	A	A	C	C	C	A	A	C
	18	189	A	A	G	G	G	G	G	G	G	G
	29	192	C	C	C	T	C	.	.	C	C	C
	125	224	G	G	A	A	G	G	G	T	G	G
	128	225	C	C	T	T	C	C	C	C	C	C
	155	234	G	G	A	G	G	G	G	G	G	G
	158	235	T	T	A	A	C	C	C	T	T	T
	164	237	G	G	A	A	G	G	G	G	G	G
	168	239	G	G	A	G	G	G	G	G	G	G
	212	253	G	G	A	A	G	G	G	G	G	G
	227	258	C	C	A	G	A	A	A	G	A	G
	236	261	G	G	A	G	G	G	G	G	G	G
	242	263	T	T	T	C	C	C	C	T	C	T
	248	265	T	T	G	G	G	G	G	G	G	G
	255	268	A	G	A	G	A	A	A	G	C	G
5	15	280	C	C	T	C	C	C	C	C	C	C
	38	287	C	C	T	C	C	C	C	C	C	C
	39	288	A	G	G	G	A	A	A	A	G	G
	52	292	T	T	C	C	T	T	T	T	T	T
	54	293	C	C	G	G	C	C	C	C	G	G
	59	294	T	T	A	G	T	T	T	T	T	T
	61	295	G	T	C	C	G	G	G	G	G	C
	62	295	A	A	A	T	A	A	A	A	A	A
	67	296	C	C	T	T	C	C	C	C	C	C
	68	296	T	T	T	A	T	T	T	T	T	T
	70	297	T	T	T	C	.	T	T	T	T	T
	71	297	G	G	G	T	.	G	G	G	G	A
	76	299	C	C	T	T	.	C	C	C	C	C
	77	299	T	T	C	A	.	T	T	T	T	T
	89	304	C	C	C	G	C	C	C	T	C	C
	94	306	C	C	C	T	C	C	C	C	C	C
	104	309	G	G	T	T	G	G	G	G	G	G
6	1	315	A	A	G	G	A	A	A	G	A	A
	3	316	A	A	G	G	A	A	A	G	A	T
	19	321	A	A	A	G	A	A	A	A	A	A
	32	325	A	A	G	G	A	A	A	G	A	A
	33	326	A	A	T	T	A	A	A	T	G	A
7	5	327	T	C	C	C	C	C	C	C	T	T
	6	328	G	A	G	A	C	C	C	G	G	G
	28	335	T	T	T	A	T	T	T	.	T	T
	37	338	C	C	C	T	C	C	C	C	C	C
	38	338	A	A	A	C	G	G	G	A	A	G
	44	340	T	A	A	T	T	T	T	T	T	T
8	1	342	T	.	.	C	T	T	T	.	T	T
	2	342	G	.	.	C	G	G	G	.	G	G

similarity with classical and nonclassical class I genes, there are individual "deleterious" nucleotide changes within the HLA-J sequences showing that HLA-J alleles are pseudogenes. These deleterious mutations are mostly located in exon 1 or exon 4. In exon 1, all three HLA-J alleles share the same single base pair deletion at position 4. In addition, the 59Kbd sequence has single base pair deletions in this same exon that are not found in the other alleles. The single nucleotide deletion in the leader peptide causing a shift in the sequence would probably abrogate translation in exon 2 at position 5 of the cda12 and Molt4px protein sequences. However, because of the 3-bp deletion, the 59Kbd sequence would read through exon 4, in which a 2-bp deletion would lead to termination of its translation in codon 195.

In exon 4, the cda12 and 59Kbd alleles share a 2-bp deletion in codon 192 and 1-bp deletion in codon 201, whereas Molt4px has only 1-bp deletion in codon 201.

TABLE III

Pairwise comparisons of entire gene[a]

A	HLA-H(Lcl1)	HLA-A2	HLA-J			HLA-B58	HLA-Cw1
			cda12	Molt4px	59Kbd		
HLA-H							
H(Lcl1)	—[b]	9.8	14.3	14.5	14	11.1	11.4
H(JY8)	0	10.0	13.6	13.7	13.4	11.1	13.5
H(12.4)	2.4	11.2	14.7	14.8	14.5	12.0	12.8
H(3.1.0)	0	9.7	13.5	13.6	13.3	11.0	13.4
HLA-A							
A2	9.8	—	13.3	13.4	13.1	11.1	13.5
A3	9.6	4.1	13.7	14.1	13.6	12.1	12.4
A24	9.2	4.1	13.9	14.3	13.7	11.2	12.8
A1	9.3	4.1	13.5	13.6	13.3	11.5	13.5
HLA-J							
J(cda12)	14.3	13.3	—	0.8	0.1	13.8	14.6
J(Molt4px)	14.5	13.4	0.8	—	0.7	14.0	14.8
J(59Kbd)	14.0	13.1	0.1	0.7	—	13.9	14.5
HLA-B							
B27	10	12.6	15.3	15.6	15.5	4.6	9.2
B35	10.9	12.8	15.8	16.1	15.9	1.9	9.0
B58	11.1	11.1	13.8	14.0	13.9	—	9.9
B57	10.6	12.8	15.6	15.9	15.7	1.6	9.5
HLA-C							
Cw1	11.4	13.5	14.6	14.8	14.5	9.9	—
Cw2	12.9	14.8	15.5	15.7	15.5	11.0	4.5
Cw3	13.4	14.3	15.4	12.3	15.4	12.5	7.2
HLA-E	17.9	18.0	19.8	19.9	19.8	18.0	17.2
HLA-F	15.6	16.6	17.2	17.2	17.2	16.1	18.1
HLA-G	13.7	13.5	14.1	14.3	14.2	13.5	14.5

B	HLA-H(Lcl1)	HLA-A2	HLA-J			HLA-B58	HLA-Cw1
			cda12	Molt4px	59Kbd		
HLA-H							
H(Lcl1)	—	7	11.4	11.3	11.3	17.4	17.3
H(JY8)	0.2	6.4	11.2	11.2	11.1	16.7	17.3
H(12.4)	2.2	7.1	12.4	12.4	12.2	17.5	17.8
H(3.1.0)	0.2	6.7	11.4	11.4	11.3	16.7	17.0
HLA-A							
A2	7.0	—	11.4	11.3	11.2	17.5	18.1
A3	7.8	4.0	11.8	11.7	11.7	17.4	18.3
A24	7.1	3.8	11.7	11.7	11.4	17.2	18.1
A1	6.7	3.0	10.9	10.8	10.7	16.7	17.6
HLA-J							
J(cda12)	11.4	11.4	—	0.5	0.2	17.7	18.0
J(Molt4px)	11.3	11.3	0.5	—	0.6	17.5	17.6
J(59Kbd)	11.3	11.2	0.2	0.6	—	16.9	17.0
HLA-B							
B27	17.4	17.5	17.7	17.5	17.0	1.4	8.8
B35	16.5	17.5	17.7	17.6	17.0	0.9	9.0
B58	17.4	17.5	17.7	17.5	16.9	—	9.0
B57	17.2	17.7	17.6	17.5	17.0	1.4	9.0
HLA-C							
Cw1	17.3	18.1	18.0	17.6	17.0	9.0	—
Cw2	17.3	18.3	18.0	17.6	17.0	8.7	1.5
Cw3	18.8	18.7	18.6	18.1	17.0	9.2	1.4
HLA-E	18.8	20.2	20.7	20.8	20.8	19.8	19.8
HLA-F	15.0	14.4	14.8	15.1	14.8	18.4	18.5
HLA-G	10.5	11.1	9.4	9.2	9.4	16.4	16.4

[a] The values indicate the percent divergence for each pairwise comparison. The percent divergence is calculated using the following formula:

$$\% \text{ Divergence} = \frac{\text{Nucleotide differences}}{\text{Shortest sequence length (bp)}} \times 100$$

A, 5′ part of HLA class I genes (5′ flanking region through exon 3); B, 3′ part of HLA class I genes (intron 3 through 3′ UTR). Sequences were obtained by Malissen et al. H(12.4) (27), Duceman and Wang H(JY8) (28), Chorney et al. H(3.1.0) (8), N'Guyen et al. (A24) (29), Weiss et al. (B27) (30), and Chertkoff et al. (B35.2) (31). The choice of HLA-A,B,C alleles compared was dictated by the availability of complete gene sequences.

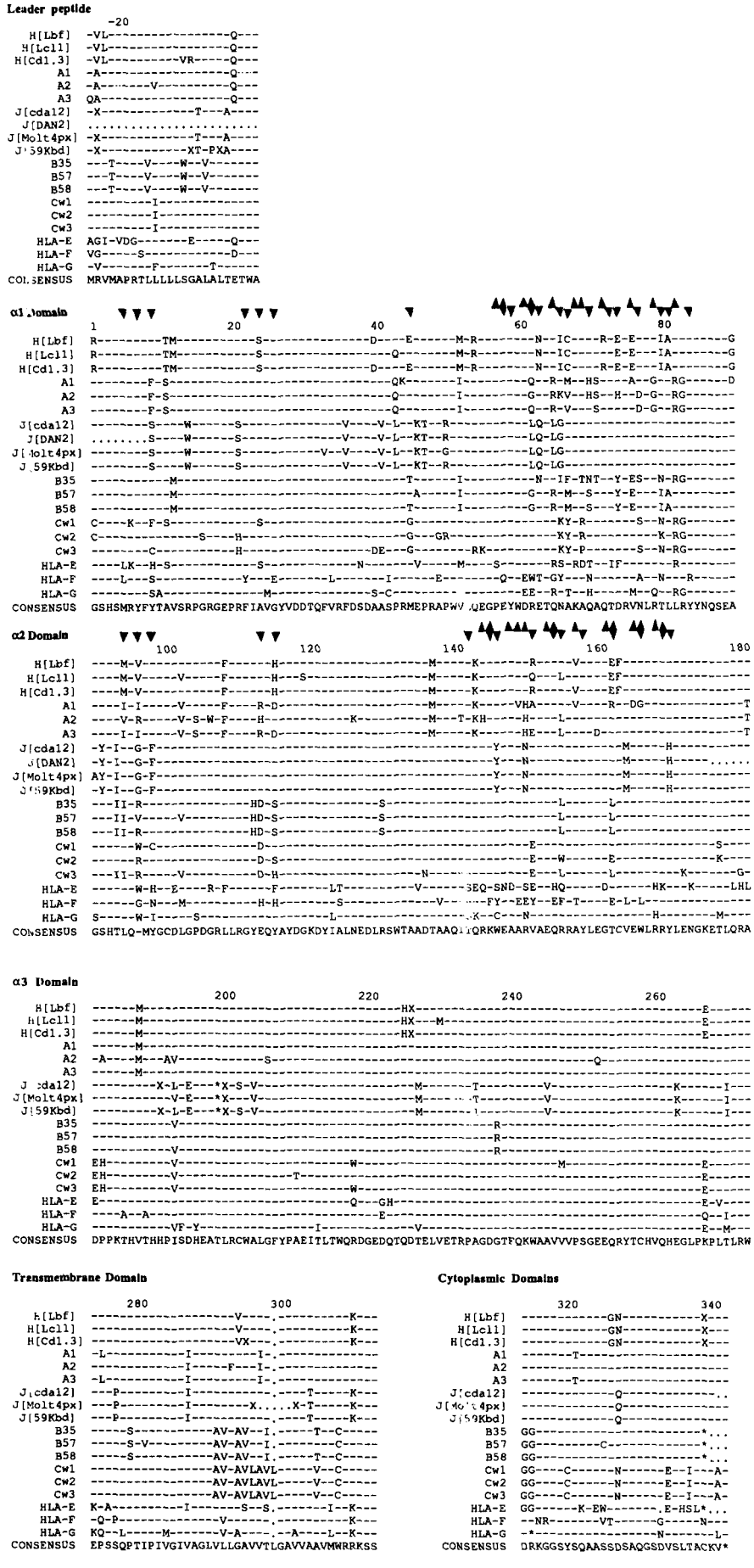

Figure 4. Comparison of the predicted protein sequences for HLA-A, B, C, E, F, G, H, and J H chains. For HLA-H and J, deletions are indicated by X. The shifts in reading frame produced by these deletions are ignored for the sake of sequence comparison. Functional positions of the peptide-binding groove are indicated by *arrowheads*. *Asterisks* indicate termination codons.

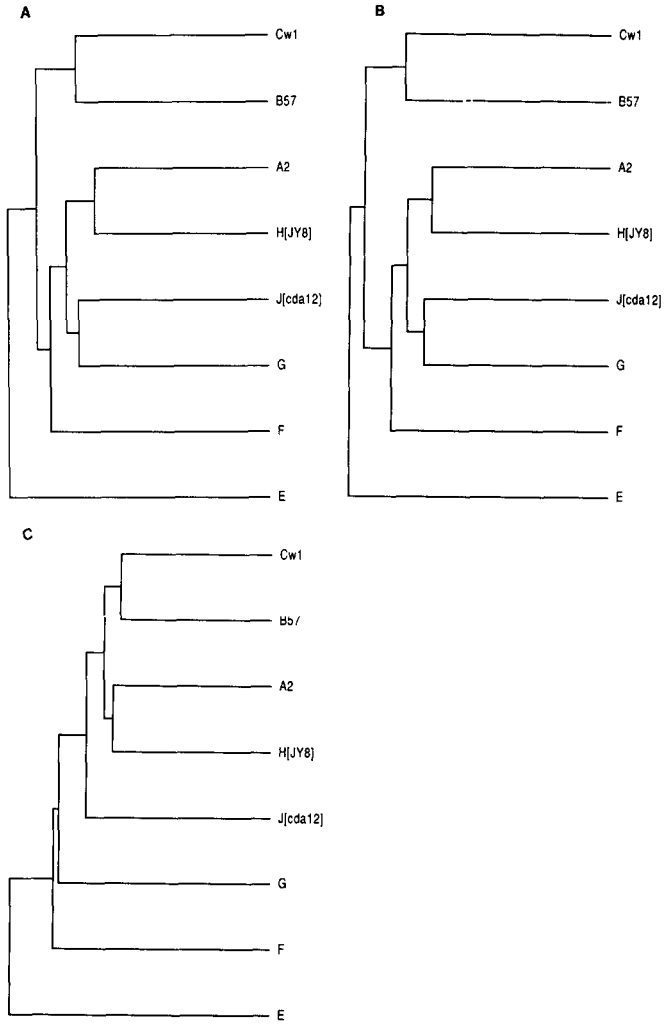

Figure 5. Phylogenetic trees showing relationships between classical and nonclassical loci using UPGMA method (14), based on the whole gene (*A*) and the 3′ part (*B*) or the 5′ part of entire HLA class I genes (*C*).

Furthermore, individual mutations are found in two of the three HLA-J alleles: cda12 has a nucleotide insertion in exon 4, and Molt4px has a deletion of 15 bp in exon 5. Clearly the HLA-J alleles cannot produce active class I Ag-presenting molecules. If the frameshift within these sequences is ignored a further distinction between Molt4px and the other HLA-J alleles is its retention of cysteine 203, which forms the disulfide bond of the α_3 domain; in both cda12 and 59Kbd the cysteine is replaced by serine, a change that might also be predicted to be deleterious in that it would prevent disulfide bond formation and thus proper folding and assembly of the molecule. That different HLA-J alleles have independently accumulated several potentially disruptive mutations is consistent with this locus being nonfunctional.

The modest polymorphism seen in the various deletions is also reflected in the point substitutions that distinguish the HLA-J alleles. The cda12 allele only differs from 59Kbd by a single substitution and three insertion/deletions and the maximum divergence seen between Molt4px and 59Kbd is still only eight substitutions and 4 insertion/deletions. This lack of polymorphism for HLA-J compared with HLA-A, B, C has also been seen for the other nonclassical class I HLA genes and pseudogenes examined (4, 7).

TABLE IV

Shared deletion/insertions among class I HLA loci

HLA Locus	Insertion/Deletion Events Shared with HLA-J	Grouping of HLA Loci	
		Combination	Occurrence
HLA-A	8	A,J	
		A,E,F,G,H,J	2
		A,G,H,J	3
		A,F,G,H,J	2
HLA-B	3	B,C,J,F	
		B,C,G,J	
		B,C,E,F,G,J	
HLA-C	4	B,C,J,F	
		C,E,F,J,G	
		B,C,G,J	
		B,C,E,F,G,J	
HLA-E	4	A,E,F,G,H,J	2
		C,E,F,J,G	
		B,C,E,F,G,J	
HLA-F	7	B,C,J,F	
		A,E,F,G,H,J	2
		C,E,F,J,G	
		B,C,E,F,G,J	
		A,F,G,H,J	2
HLA-G	9	A,E,F,G,H,J	2
		A,G,H,J	3
		C,E,F,G	
		B,C,G,J	
		B,C,E,F,G,J	
		A,F,G,H,J	2
HLA-H	7	A,E,F,G,H,J	2
		A,G,H,J	3
		A,F,G,H,J	2

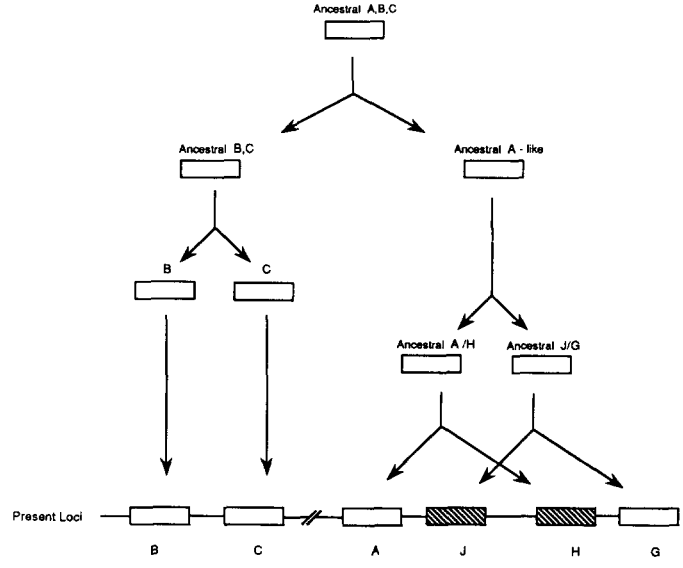

Figure 6. Schematic representation of the hypothetical evolution of the HLA-A-related loci. This representation takes into account the map position of these genes on the chromosome. Pseudogenes are indicated by *shaded boxes.*

Predicted Protein Sequences. Ignoring the shifts in reading frame caused by the various insertion/deletions, one finds that the predicted protein sequences for HLA-J alleles share many features with the classical class I genes. These include the glycosylation site at position 86 and three of the four cysteines in the extracellular domains. Of 182 residues in the α_1 and α_2 domains of class I H chains, 113 are conserved in all HLA-A and B H chains. Of these, 103 are also found in the HLA-J se-

quences. This degree of similarity argues that HLA-J was derived from an ancestral functional Ag-presenting locus and that at some point in its evolution became inactive.

As expected from the nucleotide sequences, there are few amino acid substitutions between the products of different HLA-J alleles. Those encoded by DAN2, cda12, and 59 Kbd have identical sequences and the Molt4px sequence differs at two positions (residues 33 and 50) in α_1, one (residue 91) in α_2 and three (residues 192, 194, and 203) in α_3. In contrast to the pattern observed with HLA-A and B proteins, none of the HLA-J differences are at functional positions that affect interactions with bound peptide and the TCR. An analogous pattern of substitutions was also seen for the HLA-H pseudogene (7). Thus in regions where peptide and TCR interacting residues localize, such as the helix of α_1 and the β pleated sheet of α_2, the products of HLA-J alleles exhibit no sequence diversity.

Insertion/Deletions in Class I HLA Genes. Alignment of the complete sequences for the different class I HLA genes requires the insertion of 38 gaps: 7 in the 5′ untranslated region, 1 in intron 1, 3 in intron 2, 7 in intron 3, 3 in exon 4, 3 in intron 4, 2 in exon 5, 6 in intron 5, 1 in intron 6, 1 in exon 7, 1 in intron 7, and 3 in the 3′ untranslated region. These insertion/deletions are markers that provide another way of assessing similarities between the loci (Table IV). At 12 of the gap positions the class I HLA loci divide into two groups with at least two loci in the smaller group in each case. At all but one of these positions, HLA-B and C group together and are distinct from HLA-A. HLA-J groups most frequently with HLA-A, F, G, and H. However, there are three positions where HLA-J groups with HLA-B and C and not with HLA-A. Thus, although HLA-J is overall more closely related to HLA-A than to HLA-B or C, it does share some features with HLA-B and C. This could represent remnants of sequence derived from the common ancestral gene or the result of subsequent intergenic conversion events.

DISCUSSION

The properties of HLA-J are similar in certain respects to those of HLA-H: they are both pseudogenes that map close to HLA-A and have greater sequence similarity with HLA-A than with either HLA-B or C. Whereas HLA-H is more closely related to HLA-A, HLA-J is paired with HLA-G: a nonclassical class I gene expressed in the villous trophoblast (32, 33). Sequence comparisons show HLA-A, G, H, and J represent a family of "A-related" class I loci, distinct from HLA-B and C and that share a more recent common ancestor. A scheme for the evolution of these loci, consistent with our analyses, is that initial duplication of the ancestral locus was followed by a period of locus diversification. One of these loci then became the progenitor for HLA-A and H, the other for HLA-G and J, in a subsequent duplication of the duplicated loci. Alternatively, the production of A/H and G/J may have been caused by two independent duplications in this latter stage of the evolution (Fig. 6).

This scheme is consistent with the position of these A-related genes on the chromosome. HLA-A and HLA-J are found within a distance of 50 kb whereas HLA-G maps to within 220 kb of HLA-A (3) and HLA-H within 200 kb of HLA-A (25). In addition, analysis of the locus-specific positions also support these findings. HLA-A and HLA-H share identical nucleotides at 46 positions whereas HLA-J and HLA-G are identical at 45 positions. A similar range of nucleotide differences is found by individually comparing the HLA-A-related loci to HLA-B. Thus 22, 24, and 24 nucleotides, respectively, are shared by the H-B, J-B, and G-B pairs of loci, suggesting the A-related loci are equally divergent from HLA-B.

HLA-J is a class I H chain gene that does not encode a conventional membrane-associated, Ag-presenting class I molecule. This is caused by substitutions causing translation termination in the α_3 domain. Apart from these differences, the HLA-J gene has an overall structure that closely resembles that of Ag-presenting class I H chain genes. In contrast to such genes, there appears to be little HLA-J polymorphism and that found is not at positions that directly influence Ag presentation. These properties also found for HLA-H are consistent with HLA-J being either a nonfunctional class I pseudogene or a class I gene that has evolved to perform a novel and unknown function. We favor the pseudogene interpretation, because of the similarities in sequence with Ag-presenting genes and the differences in translation termination between HLA-J alleles: it seems unlikely that a protein product of 29 amino acids would perform a novel function. The pattern of sequence differences between HLA-J alleles appears more compatible with an evolutionary pathway that led to loss of Ag-presenting function rather than one involving positive selection for a new function. However, the latter possibility cannot be ruled out for HLA-J. On available evidence this scenario appears more appropriate to HLA-G, an expressed monomorphic locus with a distinctive tissue distribution (32, 33).

HLA-H and J may have been inactivated during the duplication of the ancestral Ag-presenting locus. If so, then each duplication led to one expressed gene—A and G—and one pseudogene—H and J. Alternatively, one or both of the pseudogenes (J and H) were active after the duplication that formed them and were inactivated during subsequent evolution.

Acknowledgments. We thank Patricia Massard and Victoria Ogbuehi for preparation of the manuscript.

REFERENCES

1. **Koller, B. H., D. Geraghty, H. T. Orr, Y. Shimizu, and R. DeMars.** 1987. Organization of the human class I major histocompatibility complex genes. *Immunol. Res. 6:1.*
2. **Koller, B. H., D. E. Geraghty, R. DeMars, L. Duvick, S. S. Rich, and H. T. Orr.** 1989. Chromosomal organization of the human major histocompatibility complex class I gene family. *J. Exp. Med. 169: 469.*
3. **Schmidt, C. M., and H. T. Orr.** 1991. A physical linkage map of HLA-A, -G, -7.5p, and -F. *Hum. Immunol. 31:180.*
4. **Heinrichs, H., and H. T. Orr.** 1990. HLA non-A, B, C class I genes: Their structure and expression. *Immunol. Res. 9:265.*
5. **Mizuno, S., J. A. Trapani, B. H. Koller, B. Dupont, and S. Y. Yang.** 1988. Isolation and nucleotide sequence of a cDNA clone encoding a novel HLA class I gene. *J. Immunol. 140:4024.*
6. **Shimizu, Y., D. E. Geraghty, B. H. Koller, H. T. Orr, and R. DeMars.** 1988. Transfer and expression of three cloned human non-HLA-A, B, C class I major histocompatibility genes in mutant lymphoblastoid cells. *Proc. Natl. Acad. Sci. USA 85:227.*
7. **Zemmour, J., B. H. Koller, P. D. Ennis, D. E. Geraghty, D. A. Lawlor, H. T. Orr, and P. Parham.** 1990. HLA-AR, an inactivated antigen-presenting locus related to HLA-A: implications for the evolution of the MHC. *J. Immunol. 144:3619.*
8. **Chorney, M. J., I. Sawada, G. A. Gillespie, R. Srivastava, J. Pan, and S. M. Weissman.** 1990. Transcription analysis, physical mapping, and molecular characterization of a nonclassical human leu-

kocyte antigen class I gene. *Mol. Cell. Biol. 10:243.*

9. **Ragoussis, J., K. Bloemer, H. Pohla, G. Messer, E. H. Weiss, and A. Ziegler.** 1989. A physical map including a new class I gene (cda12) of the human major histocompatibility complex (A2/B13 haplotype) derived from a monosomy 6 mutant cell line. *Genomics 4:301.*
10. **Bodmer, J. G., S. G. E. Marsh, E. D. Albert, W. F. Bodmer, B. Dupont, H. A. Erlich, B. Mach, W. R. Mayr, P. Parham, T. Sasazuki, G. M. T. Schreuder, J. L. Strominger, A. Svejgaard, and P. I. Terasaki.** 1992. Nomenclature for factors of the HLA system, 1991. *Tissue Antigens. In press.*
11. **Girdlestone, J., and C. Milstein.** 1988. Differential expression and interferon response of HLA class I genes in thymocyte lines and response variants. *Eur. J. Immunol. 18:139.*
12. **Girdlestone, J.** 1990. Nucleotide sequence of an HLA-A1 gene. *Nucleic Acids Res. 18:6701.*
13. **Geraghty, D. E., X. Wei, H. T. Orr, and B. H. Koller.** 1990. Human leukocyte antigen F (HLA-F): an expressed HLA gene composed of a class I coding sequence linked to a novel transcribed repetitive element. *J. Exp. Med. 171:1.*
14. **Sneath, P. H. A., and R. R. Sokal.** 1973. *Numerical Taxonomy.* W. H. Freeman, San Francisco, pp. 230–234.
15. **Hansen, T., G. Markussen, G. Paulsen, and E. Thorsby.** 1991. Partial DNA sequences derived from two previously unknown HLA class I genes. *Tissue Antigens 37:16.*
16. **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.*
17. **Strachan, T., R. Sodoyer, M. Damotte, and B. R. Jordan.** 1984. Complete nucleotide sequence of a functional class I HLA gene, HLA-A3: Implications for the evolution of HLA genes. *EMBO J. 3:887.*
18. **Ooba, T., H. Hayashi, S. Karaki, M. Tanabe, K. Kano, and M. Takiguchi.** 1989. The structure of *HLA-B35* suggests that it is derived from *HLA-Bw58* by two genetic mechanisms. *Immunogenetics 30:76.*
19. **Isamat, M., J. Girdlestone, and C. Milstein.** 1990. Nucleotide sequence of an HLA-Bw57 gene. *Nucleic Acids Res. 18:6702.*
20. **Ways, J. P., H. L. Coppin, and P. Parham.** 1985. The complete primary structure of HLA-Bw58. *J. Biol. Chem. 260:11924.*
21. **Güssow, D., R. S. Rein, I. Meijer, W. de Hoog, G. H. A. Seemann, F. M. Hochestenbach, and H. L. Ploegh.** 1987. Isolation, expression, and the primary structure of *HLA-Cw1* and *HLA-Cw2* genes: evolutionary aspects. *Immunogenetics 25:313.*
22. **Sodoyer, R., M. Damotte, T. L. Delovitch, 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.*
23. **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.*
24. **Geraghty, D. E., B. H. Koller, and H. T. Orr.** 1987. A human major histocompatibility complex class I gene that encodes a protein with a shortened cytoplasmic segment. *Proc. Natl. Acad. Sci. USA 84: 9145.*
25. **Pontarotti, P., G. Chimini, C. Nguyen, J. Boretto, and B. R. Jordan.** 1988. CpG islands and HTF islands in the HLA class I region: investigation of the methylation status of class I genes leads to precise physical mapping of the HLA-B and -C genes. *Nucleic Acids Res. 16:6767.*
26. **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.*
27. **Malissen, M., B. Malissen, and B. R. Jordan.** 1982. Exon/intron organization and complete nucleotide sequence of an HLA gene. *Proc. Natl. Acad. Sci. USA 79:893.*
28. **Duceman, B., and A.-M. Wang.** 1988. A novel human major histocompatibility complex class I-type DNA sequence. *Nucleic Acids Res. 16:10391.*
29. **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.*
30. **Weiss, E. H., W. Kuon, C. Dörner, M. Lang, and G. Riethmüller.** 1985. Organization, sequence and expression of the HLA-B27 gene: a molecular approach to analyze HLA and disease associations. *Immunobiology 170:367.*
31. **Chertkoff, L. P., M. Herrera, L. Fainboim, and M. L. Satz.** 1991. Complete nucleotide sequence of a genomic clone encoding HLA-B35 isolated from a Caucasian individual of Hispanic origin: identification of a new variant of HLA-B35. *Hum. Immunol. 31:153.*
32. **Ellis, S. A., M. S. Palmer, and A. J. McMichael.** 1990. Human trophoblasts and the choriocarcinoma cell line BeWo express a truncated HLA class I molecule. *J. Immunol. 144:731.*
33. **Kovats, S., E. K. Main, C. Librach, M. Stubblebine, S. J. Fisher, and R. DeMars.** 1990. A class I antigen, HLA-G, expressed in human trophoblasts. *Science 248:220.*