Identification of a Novel TAP2 Allele in a Colombian Black Population: Gene Conversion, Ancestral Intermediate, or Convergent Change?

Mary Carrington,* J. Claiborne Stephens,† Wei-Yu Ma,*† Maureen Martin,*† Anita Harding,*† Janelle Noble,‡ Henry Erlich,‡ Dean Mann,§ Cesar Arango,∥ Ramiro Jaramillo,∥ Mauricio Concha,∥ Elizabeth Maloney,# and William Blattner**

The TAP (transporter associated with antigen processing) genes encode molecules which play an integral role in immune responsiveness by mediating transport of antigenic peptides from the cytosol into the endoplasmic reticulum. Genes encoding these molecules reside in the class II region of the human major histocompatibility complex, HLA. The human TAP1 and TAP2 genes have limited polymorphism (Bodmer et al. 1995), unlike the extensive polymorphism characteristic of their classical class II neighbors DRB1, DQB1, DQA1, and DPB1. Further, current evidence indicates that the polymorphisms in the human TAP genes have no functional significance (Obst et al. 1995).

Gene conversion has played a significant role in the generation of polymorphism of the HLA class I and II genes. Direct evidence for interallelic gene conversion in the DPB1 gene using sserpin typing has been reported (Zangenberg et al. 1995), confirming conclusions of evolutionary analyses. It does not necessarily follow that mechanisms involved in generation of diversity occur at a particularly high rate within genes such as these, but, rather, that natural selection promotes retention of polymorphism when it has functional significance. It is possible that gene conversion also occurs in conserved genes, such as TAP1 and TAP2, but generally goes undetected due to the limited polymorphism characterizing these genes. In addition, since the TAP polymorphisms have no apparent functional significance, selection favoring new variants would not be expected.

The last exon of the TAP2 gene, exon 11, contains three well-characterized polymorphic sites (Bodmer et al. 1995) which are located within 98 bp of one another; G or A at cDNA position 1993, C or T at position 2059, and T or G at position 2091, each of which are nonsynonymous changes. Extensive TAP2 genotyping of well over a thousand Caucasians in our laboratory and others (Colonna et al. 1992; Powis et al. 1992) has shown only recently on a + + + background. All alleles were inherited in a Mendelian fashion.

If the novel + - + allele arose recently from a + + + background, then it is reasonable to expect that flanking polymorphic sites would be typical of those associated with the + + + allele. On the other hand, if the + - + allele is ancestral to either standard allele, then it is likely that the flanking polymorphisms would also be intermediate between the two standard alleles. Polymorphisms in flanking regions including segments within both intron 10 and the 3′ untranslated region were identified by SSCP analysis of DNA from a panel of 16 B lymphoblastoid homozgyous typing cell lines (HTCs; Yang et al. 1989) known to be homozgyous throughout most of the 4-Mb region of HLA. Eight of the HTCs are homozgyous for the + + + allele and eight are homozgyous for the + - - allele, which facilitates determination of phase of flanking polymorphic sites. All samples were different from all + + + samples in both flanking regions, and there were additional differences within each of these two groups. In all five cases, the novel + - + allele had banding patterns (in the exon 11 flanking regions) that were identical to those for a specific + + + variant. Phase was determined by segregation analysis in the families.

Sequence analysis of each of the SSCP-defined polymorphic segments was performed on at least one HTC representing each of the unique banding patterns and on one Colombian individual with the genotype + + +/+ + - . Samples with identical SSCP banding patterns in a given region were assumed to have identical sequences in that region. Sixteen variable sites were identified, including two insertion/deletion (indel) polymorphisms (fig. 1A). As expected for a short region (<900 bp), linkage disequilibrium appears to be strong, and there were only five distinct sequences in our sample of 16 HTCs (A, B, C, E, and F, fig. 1A). Most of the variants (10/16, including a 10-bp indel) differentiated the + + + sequence types from the + - - types in all HTC samples tested. Two distinct sequences carried
The frequency of the novel allele in 135 independent chromosomes from the 25 Colombian families was 3.7%. Samples from West African Blacks and North American Blacks were typed at exon 11 of TAP2 to determine whether the novel allele could have arisen in the Colombian Black population or was older and more widely spread among other populations of African origin. The novel allele was identified in 8/236 (3.4%) chromosomes from West African Black and 31/856 (3.6%) chromosomes from North American Black individuals, all of which were unrelated. This observation is consistent with the view that the $++-$ allele is of more recent origin than the other sequence variants and suggests an origin in the range of 10,000 to 100,000 years ago given its restriction to populations of African origin.

 Colombian families in which the novel allele was observed were also typed for variations in DRB1, DQA1, DQB1, TAP1, and DPB1 (for typing methods see Bugawan, Begovich, and Erlich 1990; Erlich et al. 1991). In five of four families, DNA from the father was not available for analysis, but DNA from the mother as well as one to five siblings was typed allowing inference of one or both paternal haplotypes. The novel allele was paternally derived in all five families, assuming that in no case did a novel mutation occur in the maternal germ line. Four different class II haplotypes (spanning a region of about 500 kb) containing the novel allele were observed (table 1). We have previously shown disequilibrium between the standard TAP2 alleles and both DQB1 and DPB1 (Carrington et al. 1994; Klitz et al. 1995). Therefore, if the novel allele is extremely new, it should be in strong linkage disequilibrium with variants on both sides. Our finding (table 1), while suggestive of linkage disequilibrium, shows that several recombination events have occurred since the origin of the new allele. It is likely that different copies of the $++-$ allele are identical by descent, since if site 2059 were a hot spot for mutation or gene conversion, the novel allele would also have been detected in Caucasian populations.

Most molecular evidence points to greater genetic diversity among African populations than among non-African populations (reviewed in von Haeseler, Sajantila, and Pääbo 1995). Therefore, if the ancestral intermediates between $+++$ and $---$ are still present in humans, African populations would be a logical place to look. If more extensive sampling does reveal additional intermediate sequences, we could refine the phylogenetic locations of the 10 changes on the branch between $---$-like sequences and $+++$-like sequences (fig. 1B). The novel TAP2 allele would appear to be an excellent marker for human migration patterns, and the level of divergence (1.5%) between $+++$ and $---$
sequences makes the natural history of this locus an intriguing object for further study.

LITERATURE CITED


CHARLES F. AQUADRO, reviewing editor

Accepted May 10, 1997

Table 1
Class II Haplotypes from Individuals Carrying the Novel TAP2 Allele

<table>
<thead>
<tr>
<th></th>
<th>DRB1</th>
<th>DQA1</th>
<th>DQB1</th>
<th>TAP2</th>
<th>TAP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>379*</td>
<td>665*</td>
<td>687*</td>
<td>333*</td>
<td>637*</td>
</tr>
<tr>
<td>1 ...</td>
<td>0301</td>
<td>0501</td>
<td>0201</td>
<td>Val</td>
<td>Thr</td>
</tr>
<tr>
<td>2 ...</td>
<td>0701</td>
<td>0201</td>
<td>0201</td>
<td>Thr</td>
<td>Gln</td>
</tr>
<tr>
<td>3 ...</td>
<td>1503</td>
<td>0102</td>
<td>0602</td>
<td>Thr</td>
<td>GlY</td>
</tr>
<tr>
<td>4 ...</td>
<td>0301</td>
<td>0501</td>
<td>0701</td>
<td>Thr</td>
<td>GlY</td>
</tr>
<tr>
<td>5 ...</td>
<td>0301</td>
<td>0501</td>
<td>0201</td>
<td>Thr</td>
<td>GlY</td>
</tr>
</tbody>
</table>

*Codon positions. Positions 379, 665, and 687 of TAP2 correspond with variable sites at cDNA positions 1993, 2059, and 2091, respectively.