Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region

Dagan A. Loisel, Matthew V. Rockman, Gregory A. Wray, Jeanne Altmann, and Susan C. Alberts

This information is current as of October 2006.

Supplementary Material	Supplementary material can be found at: www.pnas.org/cgi/content/full/0607662103/DC1
	This article has been cited by other articles: www.pnas.org#otherarticles
E-mail Alerts	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.
Rights & Permissions	To reproduce this article in part (figures, tables) or in entirety, see: www.pnas.org/misc/rightperm.shtml
Reprints	To order reprints, see: www.pnas.org/misc/reprints.shtml

Notes:

Ancient polymorphism and functional variation in the primate MHC-*DQA1* 5′ cis-regulatory region

Dagan A. Loisel*†, Matthew V. Rockman‡§, Gregory A. Wray*¶, Jeanne Altmann†§‖**, and Susan C. Alberts*·**

*Department of Biology, Duke University, Durham, NC 27705; ‡Lewis–Sigler Institute for Integrative Genomics and §Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544; ¶Institute for Genome Sciences and Policy, Duke University, Durham, NC 27705; ‖Department of Conservation Biology, Chicago Zoological Society, Brookfield, IL 60513; and **Institute of Primate Research, National Museums of Kenya, Nairobi, Kenya

Precise regulation of MHC gene expression is critical to vertebrate immune surveillance and response. Polymorphisms in the 5′ proximal promoter region of the human class II gene *HLA-DQA1* have been shown to influence its transcriptional regulation and may contribute to the pathogenesis of autoimmune diseases. We investigated the evolutionary history of this cis-regulatory region by sequencing the *DQA1* 5′ proximal promoter region in eight nonhuman primate species. We observed unexpectedly high levels of sequence variation and multiple strong signatures of balancing selection in this region. Specifically, the considerable *DQA1* promoter region diversity was characterized by abundant shared (or trans-species) polymorphism and a pronounced lack of fixed differences between species. The majority of transcription factor binding sites in the *DQA1* promoter region were polymorphic within species, and these binding site polymorphisms were commonly shared among multiple species despite evidence for negative selection eliminating a significant fraction of binding site mutations. We assessed the functional consequences of intraspecific promoter region diversity using a cell line-based reporter assay and detected significant differences among baboon *DQA1* promoter haplotypes in their ability to drive transcription *in vitro*. The functional differentiation of baboon promoter haplotypes, together with the significant deviations from neutral sequence evolution, suggests a role for balancing selection in the evolution of *DQA1* transcriptional regulation in primates.

balancing selection | cis-regulatory evolution | major histocompatibility complex | primate molecular evolution

Changes in cis-regulatory DNA sequence can alter the timing, level, and tissue distribution of gene expression, potentially modifying the emergence and expression of key phenotypic traits (1, 2). However, despite its biological importance, relatively little is known about the nature and patterns of extant cis-regulatory variation, the evolutionary forces acting on such variation, or the potential functional consequences of that diversity.

The MHC is a multigene family essential to the development and activation of vertebrate adaptive immunity. Within the MHC, the class II region contains the classical genes that encode the distinct α (A) and β (B) subunits that constitute functional class II molecules (in humans, HLA-DR, HLA-DQ, and HLA-DP). A primary function of these heterodimeric cell-surface molecules is to present short peptides from extracellular pathogens to helper T cells, an interaction that is essential for the initiation and regulation of adaptive immune responses against pathogens (3). Because of this, the class II region is genetically associated with more diseases than any other comparably sized region in the human genome (4). Furthermore, there is substantial evidence, across multiple taxa, that class II gene coding sequence has evolved in response to natural selection (5–7).

Cell surface expression of class II molecules is finely and tightly regulated to balance an appropriately vigorous response to pathogens with minimal collateral damage to self tissues (4). Constitutive expression is normally restricted to specialized antigen presenting cells, e.g., thymic epithelial cells, macro-phages, and B cells, but expression on other cell types can be induced by immune activators (3, 8). Class II expression is also developmentally regulated during immune cell maturation (3, 8). Precise regulation of class II expression is essential for normal immune function; defects in proper regulation can have profound immune consequences (9, 10).

Expression of class II molecules is regulated primarily at the level of transcription (3, 4, 10). Conserved cis-regulatory sequence elements in the 5′ proximal promoter region, i.e., the 300-bp region immediately upstream of translation start, direct most aspects of constitutive and inducible expression in class II genes (3, 8, 11). These elements are short sequences, typically 7–14 bp, that function as transcription factor binding sites. Additional gene-specific cis-regulatory elements further affect expression phenotypes (12) and may provide additional levels of regulatory control. Overall, the proximal promoter region behaves as a single functional unit, with the sequence, spacing, and orientation of the regulatory elements synergistically contributing to the expression phenotype (10).

Mutations in the proximal promoter region of the human class II gene *HLA-DQA1* (Online Mendelian Inheritance in Man database accession no. 146880) significantly affect constitutive and inducible expression in reporter assays (13). Furthermore, *DQA1* promoter region polymorphisms are associated with quantitative expression differences in cell lines (14), and they affect the strength of transcription factor binding (15). Moreover, specific human *DQA1* proximal promoter alleles may be associated with increased susceptibility to a number of autoimmune and infectious diseases (16–18).

The overall importance of proper immune regulation, the observed population genetic diversity of the *DQA1* promoter in humans (19–21), and the demonstrated functional significance of this diversity make the proximal promoter region of *DQA1* a likely target of natural selection (22). To explore this possibility, we investigated the evolutionary history of the *DQA1* proximal promoter region in primates and characterized cis-regulatory variation at both the nucleotide and functional levels. To assess the functional importance of observed diversity, we used a cell line-based reporter assay to quantify expression differences between *DQA1* promoter region haplotypes observed in the wild Amboseli baboon population (23).

Results

Extensive Polymorphism but Not Fixed Divergence Characterizes the *DQA1* Promoter Region in Primates. We sequenced the 270-bp region immediately upstream of the *DQA1* translation start site

EVOLUTION
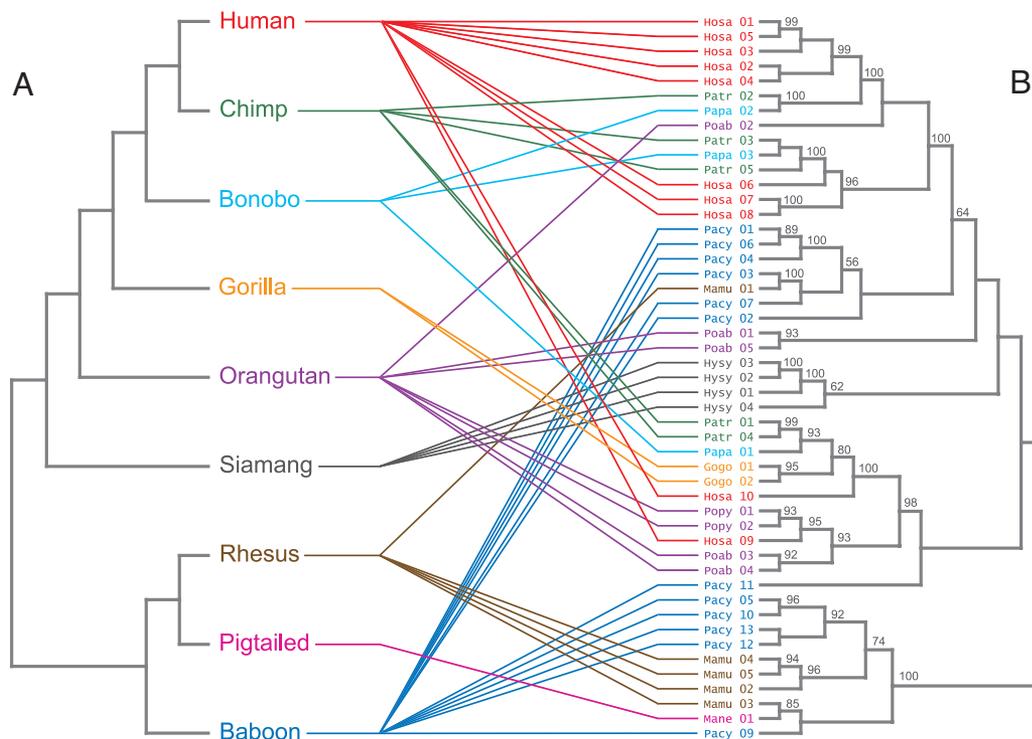
Fig. 1. Incongruity of primate species tree and *DQA1* promoter region gene tree. (*A*) Species tree of the nine species from which *DQA1* promoter sequences were obtained. (*B*) The most probable *DQA1* promoter tree topology from the Bayesian analysis, with posterior probabilities of branch support. Branch lengths for both trees are arbitrary. Species and gene tree were reconciled by using Page's method (38) for reconciling phylogenetic trees (69).

(i.e., the *DQA1* proximal promoter region) in 80 individuals from eight nonhuman primate species (*n* = 10 chimpanzees, 3 bonobos, 1 gorilla, 5 orangutans, 3 siamangs, 53 baboons, 4 rhesus macaques, and 1 pigtailed macaque) (Table 1, which is published as supporting information on the PNAS web site). Combining the 10 published human sequences (19–21) with our sequencing results, we identified 47 unique *DQA1* proximal promoter region haplotypes across the nine primate species (Fig. 4, which is published as supporting information on the PNAS web site). We observed 125 mutations, occurring at 113 nucleotide sites in these 47 haplotypes, yielding an average density of one mutation every 2.2 bp. The proportion of sites that varied within or between species (≈42%) and the mean number of nucleotide differences per site between the 47 unique haplotypes (0.071) were considerable, in contrast to previous descriptions of primate class II promoter region diversity (24, 25).

The extensive allelic and nucleotide diversity we observed in the *DQA1* promoter region was characterized by a profound lack of fixed divergence between species. For example, we identified 34 mutations in the 10 human *DQA1* promoter haplotypes and 26 mutations in the five chimpanzee haplotypes but found no fixed differences between the two species (Table 2, which is published as supporting information on the PNAS web site). In contrast, recent analyses of the initial chimpanzee genome sequence estimated that fixed differences accounted for 78–86% of the genome-wide divergence between humans and chimpanzees (26). The lack of fixed divergence in the *DQA1* promoter region is also apparent in comparisons of more distantly related primate species. Remarkably, we found no fixed differences between human haplotypes and those from three Old World monkey species (baboon, rhesus, and pigtailed macaque) (Table 2). This is unexpected given that the four species last shared a common ancestor ≈25 million years ago (27, 28). In fact, the only fixed differences we observed involved species for which we had

extremely small sample sizes (siamang, *n* = 2 unrelated individuals; gorilla, *n* = 1) and thus may be due to sampling effects.

Under the neutral model, levels of divergence between species and polymorphism within species should be correlated, because both are related to the neutral mutation rate (29). The lack of fixed differences in the *DQA1* promoter region, particularly in light of the substantial polymorphism observed in these species, therefore suggests balancing selection. Because it maintains allelic lineages for long periods of time, balancing selection tends to enhance levels of polymorphism relative to divergence (29). However, nonselective forces may also influence the ratio of polymorphism to divergence (30).

Primate *DQA1* Promoter Haplotypes Exhibit Widespread Trans-Species Polymorphism. The elevated diversity of the *DQA1* promoter region in primates may be due to long-term maintenance of ancestral polymorphism or to diversification after species divergence. To differentiate between these possibilities, we compared promoter sequences from the nine primate species (Fig. 1*A*) and used Bayesian methods to infer the phylogenetic relationships of the *DQA1* promoter region haplotypes (Fig. 1*B*). The most probable *DQA1* promoter tree shows several well supported clusters consisting of haplotypes from different species (Fig. 1*B*). This extensive trans-species polymorphism is highly unlikely under neutrality and is expected only after long-term maintenance of allelic lineages, e.g., across speciation events (31–33). Thus, the occurrence of trans-specificity provides strong evidence of long-term balancing selection (32–34). Although it is frequently observed in phylogenies of MHC exon 2 coding sequence (35), the trans-species polymorphism we observe in the *DQA1* promoter region is one of the first examples of trans-specificity in a primate cis-regulatory region (36, 37).

To quantify the degree of trans-specificity observed in the primate *DQA1* promoter tree, we estimated the minimum number of "deep coalescent" events necessary to reconcile a gene

Transcription factor binding sites and motifs

	W	NF-κB	S	J	X1	X2	Y
	A. GCTGCATC B. -----G-- C. A-------	A. GGGGATTGCCC B. -------T---	A. AGGGCTC B. G------ C. ------T	A. CTTCAACTAG B. -----G--- C. ---A------ D. ----C-----	A. ACTAGTAACTGAGG B. -------------A C. G------------A D. -----------AA E. -----------C- F. -----------A- G. --------T---- H. --------T--A-	A. GGTGTCA B. -A----- C. AA----- D. C------ E. A------	A. CTAATTGGCC B. ------A--- C. T-----A--- D. --------T- E. --G------- F. -----C----
Human	A,B	A,B	A	A,B	A,B,C	A,B	A,B
Chimpanzee	A,B	A,B	A	A	B,D,E	B,C,D	A,B
Bonobo	A,B	A,B	A	A	A,B,D	A,B,C	A,B
Gorilla	B	B	A	A	A	A	B
Orangutan	A	A,B	A	A	A,B	A,B	A,B,C,D
Siamang	A	A,B	B,C	A	A,F	A,B	A,D
Baboon	A,C	A	A,B,C	A,C	A,G	A	A,D,E,F
Rhesus macaque	A,C	A	A	A,D	A,G,H	A,E	D,E
Pigtailed macaque	A	A	A	A	G	A	E

25 18 14 7 6 2 MYA

Fig. 2. Transcription factor binding site motifs in the *DQA1* proximal promoter in primates. Letters A–H refer to unique transcription factor binding site motifs, with the corresponding sequence represented above. W, NF-κB, S, J, X1, X2, and Y refer to binding sites identified in the *DQA1* promoter (3, 11, 12). Branch lengths of the species phylogeny on the left (27, 28) are roughly proportional to the timeline indicated below it. MYA, million years ago.

tree with the accepted species tree (38). A deep coalescent event, in this case, describes an interspecific coalescence of alleles. We compared the most probable *DQA1* promoter tree from the Bayesian analysis (Fig. 1*B*) to the primate species tree (Fig. 1*A*) and found that 37 deep coalescent events are required to account for the incongruence between the trees. Because there is uncertainty in the phylogenetic reconstruction of any single tree, we analyzed the 100 most frequently sampled *DQA1* promoter topologies from the Bayesian analysis and obtained a mean of 37.8 deep coalescent events (95% credible interval of 32–44 events). Under the simplest models of neutrality and drift, there should be complete congruence between gene and species trees, and a set of species separated by $>4 N_e$ generations should have $(n - 1)$ interspecific coalescent events, where n is the number of species sampled (7, 39). Therefore, the mean of 37.8 interspecific coalescent events observed in our *DQA1* analysis is nearly 5-fold higher than that expected when there is complete congruence between the gene and species trees ($n - 1 = 8$). Our estimate should be taken with caution, however, because it does not factor in the effect of recombination, which may lead to an overestimate of the number of trans-specific lineages (40).

This pattern of trans-species polymorphism is particularly evident in pairwise species comparisons of individual mutations. For example, 18 of the 34 mutations observed in human *DQA1* promoter haplotypes are shared with chimpanzees (Table 2). Because neutral polymorphism is expected to persist for $\approx 4 N_e$ generations, the probability of observing polymorphisms shared between humans and chimpanzees, two species that diverged $\approx 20 N_e$ generations ago, is extremely low (41, 42). This prediction is supported empirically: a large-scale comparison of 6,700 chimpanzee SNPs and 10,000 human SNPs identified only eight trans-species polymorphisms (42), and a microarray analysis of 397 human SNPs failed to find any polymorphisms shared with common chimpanzees (43). Furthermore, shared polymorphisms in the *DQA1* promoter region were not restricted to closely related species like humans and chimpanzees. Multiple shared polymorphisms were observed in nearly all pairwise species comparisons (Table 2), including the comparisons of humans and Old World monkeys, species that diverged ≈ 25 million years ago (27, 28). This maintenance of substantial levels of ancestral polymorphism across multiple speciation events,

occurring at different times in primate evolution, supports the hypothesis of long-term balancing selection on the *DQA1* promoter region.

Transcription Factor Binding Sites Are Polymorphic Within and Shared Between Species. The functional importance of the cis-regulatory elements in the proximal promoter region of human MHC class II genes has been examined in great detail (reviewed in refs. 3 and 8–12). The *DQA1* proximal promoter region contains four essential transcription factor binding sites (the S box, X1 box, X2 box, and Y box) that are conserved across class II genes (3). In addition, this region contains three *DQA1*-specific binding sites [the NF-κB box (44), the W box (45), and the J box (46)] that affect *DQA1* transcription. To elucidate the evolutionary history of these cis-regulatory elements, we analyzed patterns of intraspecific and interspecific diversity in the seven binding sites (Fig. 2).

There was considerable intraspecific diversity in transcription factor binding site sequence in the *DQA1* promoter region. In the seven species represented by more than one individual (humans, chimpanzees, bonobos, orangutans, siamangs, baboons, and rhesus macaques), most binding sites were polymorphic; on average, five of seven binding sites were variable within a given species (Fig. 2). Most binding sites were dimorphic, but there were many examples of three or four different sequence motifs segregating at a single binding site. Considering just the baboon haplotypes, for example, there were two invariant binding sites (NF-κB and X2), three dimorphic binding sites (W, J, and X1), one binding site polymorphic for three motifs (S), and one binding site polymorphic for four motifs (Y) (see Fig. 2). Given the empirically documented importance of binding site sequence in transcriptional regulation, this segregating diversity has the potential to seriously affect the regulation of *DQA1* expression.

Transcription factor binding site polymorphisms were often shared among several species (Fig. 2). In some cases, sharing extended across tens of millions of years of primate evolution. For example, the same two Y box motifs (A and B in Fig. 2) were found in four hominoid species (humans, chimpanzees, bonobos, and orangutans), a group that last shared a common ancestor ≈ 12–15 million years ago (27, 28). Interestingly, these two Y box motifs were shown to have different transcription factor binding affinities and
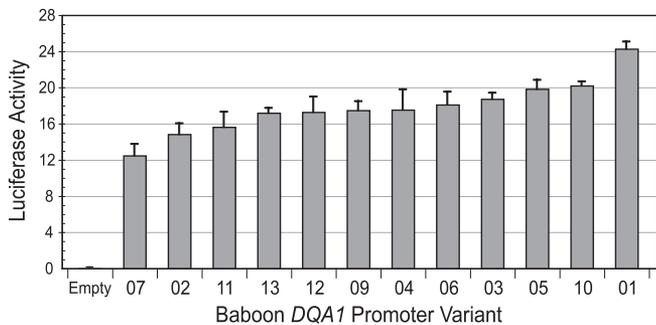
Fig. 3. Luciferase activity induced by baboon *DQA1* promoter haplotypes in a transient transfection reporter assay. Bars represent luciferase activity of 12 different baboon *DQA1* promoter constructs calculated as the ratio of firefly to *Renilla* (coreporter) luminescence. Error bars represent standard error of the mean estimated from three replicate experiments. Numbers on the *x* axis correspond to the unique baboon *DQA1* promoter haplotypes, and "Empty" refers to the promoter-less pGL3 basic reporter construct.

were associated with divergent *DQA1* expression profiles in humans (15). Even more striking is the presence of two NF-κB binding site motifs (A and B in Fig. 2) in humans, chimpanzees, bonobos, orangutans, and siamangs, a collection of species that last shared a common ancestor ≈18 million years ago (28).

Functionally important cis-regulatory sequence may be more conserved than surrounding sequence because of purifying selection (2). To test whether nucleotide diversity differs between cis-regulatory elements and intervening sequence in the *DQA1* promoter, we compared the number of mutations per nucleotide in the experimentally validated transcription factor binding sites to the number of mutations per nucleotide in adjacent nonbinding site sequence. We estimated total mutations per nucleotide because there are 12 instances of multiple mutations at a single nucleotide, the majority of which (11 of 12) occur in nonbinding site sequence. Comparing across the 47 unique *DQA1* promoter region haplotypes, the number of mutations per binding site nucleotide ($17/60 = 0.28$) is significantly lower (Fisher's exact test, $P = 0.028$) than the number of mutations per nonbinding site nucleotide ($108/210 = 0.51$). The preferential conservation of transcription factor binding sites suggests functional constraint due to purifying selection.

Baboon *DQA1* Promoter Polymorphisms Have Functional Effects on Transcription *in Vitro*. Changes in cis-regulatory sequence affect phenotype in highly variable and often unpredictable ways (2). Some changes are silent, others affect transcription factor binding and/or transcriptional output, and a small subset induce change in phenotypes related to physiology, behavior, disease resistance, and life history (47). To assess the functional consequences of mutations in cis-regulatory sequence, we used a reporter assay system to experimentally determine the effect of naturally occurring baboon *DQA1* promoter region diversity on transcriptional activation *in vitro*. We cloned each of the 12 baboon *DQA1* promoter haplotypes (size = 310 or 318 bp) upstream of a luciferase reporter gene. The resulting constructs contained all seven of the transcription factor binding sites (Fig. 2) found in the *DQA1* proximal promoter region. The 12 baboon *DQA1* promoter constructs were transiently transfected into an immortalized baboon lymphocyte cell line, and the resulting luciferase activity was quantified.

The baboon *DQA1* promoter constructs differed significantly in their ability to drive expression of the reporter gene in baboon lymphoblast cells (Fig. 3) (ANOVA, $P = 0.00017$, $F = 6.07$, df = 11). Although many of the constructs exhibited similar mean

expression levels, there was a 2-fold range of variation between the lowest (variant 07) and highest (variant 01) mean expression values (Fig. 3). This suggests that the intraspecific *DQA1* promoter diversity represented in the constructs is both visible to and differentially utilized by the transcriptional machinery of the cells.

Discussion

We have documented elevated diversity, a lack of fixed differences, considerable trans-species polymorphism, and extensive sharing of transcription factor binding site motifs in the *DQA1* 5′ proximal promoter region in primates. These results suggest nonneutral evolution and are characteristic of genes evolving under balancing selection. The demonstrated functional effect of promoter polymorphism in the reporter assay is also consistent with the action of balancing selection, inasmuch as the result provides evidence for the functional differentiation of alleles, a necessary condition for natural selection to act. The preferential conservation of transcription factor binding site sequence, however, suggests that purifying selection may be acting in addition to balancing selection, eliminating deleterious mutations. Evidence of balancing selection shaping patterns of 5′ cis-regulatory diversity has recently been documented in human population genetic studies of other immune system genes, such as the class Ib gene *HLA-G* (48), the class II genes *HLA-DPA1* and *HLA-DPB1* (49), and the cell surface chemokine *CCR5* (50). The *HLA-G* and *CCR5* promoter regions, for example, showed similar signatures of balancing selection in human populations: elevated diversity, more polymorphism relative to divergence, an excess of intermediate frequency alleles, and a deep divergence between allelic lineages (48, 50).

In addition to selection acting directly on the *DQA1* proximal promoter region, other evolutionary forces may have contributed to the observed levels and patterns of diversity. First, the evolution of the *DQA1* promoter region may have been influenced by its proximity, i.e., physical linkage, to the highly polymorphic second exon of the gene, a coding sequence located 4 kb downstream that is known to be evolving under balancing selection (7). Balancing selection on one locus increases levels of neutral variation at closely linked sites, as a function of the physical distance between the two loci, intensity of selection, recombination rate, effective population size, and mutation rate (33, 34, 51, 52). In the human genome, this genetic hitchhiking has the potential to impact neutral diversity tens of kilobases away from a selected site (33, 53) and has been proposed to explain the elevated silent site diversity (53) and increased polymorphism in a pseudogene (54) in the human MHC region. Second, epistatic selection may have generated linkage disequilibrium between nucleotides in the proximal promoter region and the second exon. In this case, the signal of hitchhiking emerges because recombination between epistatic loci produces relatively unfit haplotypes that are eliminated by selection (55). Selection against recombinants has been proposed to explain the strong, long-range linkage disequilibrium between neutral and selected sites in 100-kb haplotypes of the human MHC class II region (56). Examples of epistatic interactions between cis-regulatory variants and coding polymorphisms were quite common in a recent survey of human cis-regulatory evolution (47).

Several lines of evidence suggest that the evolution of the primate *DQA1* promoter region is not simply due to linkage disequilibrium with the second exon. First, although hitchhiking may have some influence on neutral diversity over a large chromosomal region, theory predicts that the strength of the effect decreases rapidly with increasing chromosomal distance (51, 52, 55). Likewise, empirical studies have shown that the elevated diversity and signatures of selection found in the

second exon of MHC loci decay rapidly in flanking intronic sequence [i.e., within a few hundred base pairs in human class I (57) and class II loci (58)] and neighboring exons (52). This rapid decay was also apparent in the MHC genes of fish and birds, taxa with very different demographic, selective, and recombination histories (59, 60). Second, recent models of trans-species polymorphism (and their application to empirical data sets) demonstrated that the probability of observing a trans-species polymorphism at neutral sites located >1 kb away from the selected site is very low and that regions of trans-specificity are quite short, at most a few hundred base pairs (32). This suggests that the strong signal of trans-species polymorphism we observe in the *DQA1* promoter region is not likely to be a result of hitchhiking, considering the nearly 4 kb of sequence that separates the promoter and second exon. Third, although the linkage between *DQA1* promoter and second exon is often described as strong, there is considerable evidence for recombination, both historical and recent, between these loci in human populations (19, 20). These recombination events have the potential to break down linkage disequilibrium and mitigate the hitchhiking effect (7, 52, 53, 55). Finally, *DQA1* promoter sequence variation has clear functional consequences on transcriptional regulation *in vitro*, as shown in human (13–15) and baboon (Fig. 3) experiments. Considering the strong signatures of balancing selection in the *DQA1* promoter region sequence, the demonstrated functional significance of diversity in the region, and the dire consequences of aberrant class II expression, it is unlikely that this cis-regulatory region is evolving as a neutral locus and that the patterns we report are due to linkage effects.

The model of balancing selection operating in the *DQA1* promoter region is presumably pathogen-driven and most likely involves a combination of heterozygote advantage (overdominance), frequency-dependent selection, and/or selection that varies in time or space. Of these, overdominant selection for multiple expression phenotypes has been a primary hypothesis for the maintenance of MHC class II promoter region diversity (22, 61). This is based on the idea that functional cis-regulatory diversity within an individual confers additional range and flexibility in important parameters of transcriptional output (e.g., site, timing, and responsiveness to cytokines), which facilitates a more precise, context-specific immune response. Overall, this would translate into a relative fitness benefit for the heterozygote (22, 61). The plausibility of tissue-type overdominance has been demonstrated in the *HLA-DQB1* promoter region, where the relative expression levels associated with two promoter haplotypes were reversed in different cell types (62).

The present study is among the first to demonstrate a role for balancing selection in cis-regulatory evolution in primates and adds to a growing list of studies linking cis-regulatory mutations and phenotypic variation. Future studies integrating molecular evolutionary and functional analyses will greatly enhance our understanding of cis-regulatory structure, function, and evolution.

Materials and Methods

Primate Sequences. Human *DQA1* promoter sequences came from diversity surveys of European and North American populations (19, 20, 63). Genomic DNA samples for eight nonhuman primate species were obtained from several sources (Table 1). The *DQA1* proximal promoter region, i.e., from −282 to −13 upstream of the translation start site, was PCR-amplified by using a touchdown program. Two forward primers, DQApFor (5′-CAGACATGCACACACCAGAGAA-3′) and DQApForII (5′-TGCACACACCAGAGAAGATTCC-3′), were used in separate reactions with the reverse primer, DQApRev (5′-GGATCATCYTCTTCCCAAGG-3′). PCR product was purified and sequenced directly. Heterozygous PCR product was

cloned, and plasmid DNA was isolated from multiple positive transformants. Sequences were produced by using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) and analyzed on an ABI 3700 capillary sequencer. Multiple identical clones were obtained for each putative *DQA1* promoter haplotype to reduce PCR artifact. Mendelian inheritance was confirmed in parent–offspring dyads. No more than two sequence variants were identified per individual, confirming that a single locus was amplified.

Sequence Analysis. Sequences were assembled by using Sequencher 4.2 (Gene Codes, Ann Arbor, MI) and aligned by using ClustalW (64). Estimates of *DQA1* haplotype sequence diversity were calculated by using DnaSP 4.00.6 (65). The mean number of nucleotide differences per site between haplotypes (π) was calculated by using the Jukes and Cantor correction for multiple hits per site. The model of sequence evolution that best fit the *DQA1* promoter alignment, HKY+G, was determined by using hierarchical likelihood ratio tests in Modeltest 3.07 (66). To infer phylogenetic relationships, Bayesian analysis incorporating the model structure specified by Modeltest was implemented by using MrBayes 3.0b4 (67). The analysis was run for 5 million generations with a sampling interval of 500 generations. Posterior probabilities were derived from trees sampled after the initial burn-in of 1 million generations. Two *DQA1* promoter alleles found in both chimpanzees and bonobos were included for both species only in the phylogenetic analysis. To quantify the degree of incongruity between the species tree and *DQA1* promoter phylogeny, we used the method of reconciled trees (38, 68) in GENETREE 1.0 (69).

Reporter Assay. The 12 baboon *DQA1* proximal promoter haplotypes were cloned into the pGL3 basic firefly luciferase reporter construct (Promega, Madison, WI). Eleven constructs were 310 bp long (from −303 to +7 after translation start), whereas one (Pacy 02) was 318 bp because of an 8-bp insertion at −22. Construct fidelity was verified by sequencing. For each experiment, 4.5 μg of experimental construct was transiently transfected, along with a *Renilla* luciferase coreporter (0.5 μg), into 3 million cell aliquots of the baboon lymphoblast cell line 26CB-1 (ATCC CRL-1495) using Nucleofector Kit V (AMAXA Biosystems, Gaithersburg, MD). Cells were maintained in 90% RPMI medium 1640/10% FBS according to American Type Culture Collection instructions. Cells were harvested 24 h after transfection. Firefly and *Renilla* luminescence were measured in relative light units by using the Dual-Luciferase Reporter Assay System (Promega) on a Turner Designs 20/20 luminometer. Luciferase activity for each construct was calculated as the ratio of firefly to *Renilla* luciferase luminescence for three replicate experiments. All baboon *DQA1* promoter constructs activated reporter gene expression at significantly higher levels (>2 orders of magnitude) than the pGL3 basic construct lacking upstream promoter sequence ("Empty" in Fig. 3). A one-way ANOVA was used to determine whether the mean expression values differed significantly among the 12 experimental constructs.

EVOLUTION

1. Davidson EH (2001) *Genomic Regulatory Systems: Development and Evolution* (Academic, San Diego).
2. Wray GA, Hahn MW, Abouheif E, Balhoff JP, Pizer M, Rockman MV, Romano LA (2003) *Mol Biol Evol* 20:1377–1419.
3. Rohn WM, Lee YJ, Benveniste EN (1996) *Crit Rev Immunol* 16:311–330.
4. Ting JP, Trowsdale J (2002) *Cell* 109:S21–S33.
5. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) *Crit Rev Immunol* 17:179–224.
6. Meyer D, Thomson G (2001) *Ann Hum Genet* 65:1–26.
7. Garrigan D, Hedrick PW (2003) *Evolution (Lawrence, Kans)* 57:1707–1722.
8. Krawczyk M, Reith W (2006) *Tissue Antigens* 67:183–197.
9. Grusby MJ, Glimcher LH (1995) *Annu Rev Immunol* 13:417–435.
10. Mach B, Steimle V, Martinez-Soria E, Reith W (1996) *Annu Rev Immunol* 14:301–331.
11. Glimcher LH, Kara CJ (1992) *Annu Rev Immunol* 10:13–49.
12. Benoist C, Mathis D (1990) *Annu Rev Immunol* 8:681–715.
13. Morzycka-Wroblewska E, Munshi A, Ostermayer M, Harwood JI, Kagnoff MF (1997) *Immunogenetics* 45:163–170.
14. Fernandez S, Wassmuth R, Knerr I, Frank C, Haas JP (2003) *Eur J Immunogenet* 30:141–148.
15. Indovina P, Megiorni F, Ferrante P, Apollonio I, Petronzelli F, Mazzilli MC (1998) *Hum Immunol* 59:758–767.
16. de la Paz Bettinotti M, Kolek A, Brunnler G, Haas P, Paul C, Hochberger M, Bartova A, Kimura A, Sasazuki T, Albert ED (1993) *Eur J Immunogenet* 20:339–407.
17. Haas JP, Kimura H, Truckenbrodt H, Suschke J, Sasazuki T, Volgger A, Albert ED (1995) *Tissue Antigens* 45:317–321.
18. Yao Z, Kimura A, Hartung K, Haas PJ, Volgger A, Brünnler G, Bönisch J, Albert ED (1993) *Immunogenetics* 38:421–429.
19. Petronzelli F, Kimura A, Ferrante P, Mazzilli MC (1995) *Tissue Antigens* 45:258–263.
20. Alaez C, Vazquez-Garcia MN, Gorodezky C (2001) *Genes Immun* 2:216–221.
21. Brunnler G, Haas JP, Fan LA, Petzl-Erler ML, Volgger A, Yao Z, Wassmuth R, Albert ED, Middleton D, Barboni F, *et al.* (1997) in *Genetic Diversity of HLA: Functional and Medical Implication*, ed Charron D (EDK, Sevres, France), pp 171–175.
22. Guardiola J, Maffei A, Lauster R, Mitchison NA, Accolla RS, Sartoris S (1996) *Tissue Antigens* 48:615–625.
23. Buchan JC, Alberts S, Silk JB, Altmann J (2003) *Nature* 425:179–181.
24. Gaur LK, Shewey L, Sharkey-Mathis D, Nepom GT (1994) *Transgenics* 1:267–276.
25. Gaur LK, Heise ER, Ting JPY (1992) *Immunogenetics* 35:136–139.
26. Mikkelsen TS, Hillier LW, Eichler EE, Zody MC, Jaffe DB, Yang SP, Enard W, Hellmann I, Lindblad-Toh K, Altheide TK, *et al.* (2005) *Nature* 437:69–87.
27. Glazko GV, Nei M (2003) *Mol Biol Evol* 20:424–434.
28. Wildman DE, Uddin M, Liu GZ, Grossman LI, Goodman M (2003) *Proc Natl Acad Sci USA* 100:7181–7188.
29. Kreitman M, Hudson RR (1991) *Genetics* 127:565–582.
30. Fay JC, Wyckoff GJ, Wu CI (2002) *Nature* 415:1024–1026.
31. Klein J (1987) *Hum Immunol* 19:155–162.
32. Wiuf C, Zhao K, Innan H, Nordborg M (2004) *Genetics* 168:2363–2372.
33. Charlesworth D (2006) *PLoS Genet* 2:e64.
34. Takahata N (1990) *Proc Natl Acad Sci USA* 87:2419–2423.
35. Bernatchez L, Landry C (2003) *J Evol Biol* 16:363–377.
36. Bontrop RE, Otting N, de Groot N, Doxiadis G (1999) *Immunol Rev* 167:339–350.
37. Vallejo A, Pease L (1995) *J Immunol* 154:3912–3921.
38. Page RDM, Charleston MA (1997) *Mol Phylogenet Evol* 7:231–240.
39. Edwards SV, Chesnut K, Satta Y, Wakeland EK (1997) *Genetics* 146:655–668.
40. Schierup MH, Mikkelsen AM, Hein J (2001) *Genetics* 159:1833–1844.
41. Clark AG (1997) *Proc Natl Acad Sci USA* 94:7730–7734.
42. Asthana S, Schmidt S, Sunyaev S (2005) *Trends Genet* 21:30–32.
43. Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, *et al.* (1999) *Nat Genet* 22:164–167.
44. Dedrick RL, Jones PP (1990) *Mol Cell Biol* 10:593–604.
45. Koch W, Benoist C, Mathis D (1989) *Mol Cell Biol* 9:303–311.
46. Sugawara M, Ponath PD, Shin J, Yang Z, Strominger JL (1991) *Proc Natl Acad Sci USA* 88:10347–10351.
47. Rockman MV, Wray GA (2002) *Mol Biol Evol* 19:1991–2004.
48. Tan Z, Shon AM, Ober C (2005) *Hum Mol Genet* 14:3619–3628.
49. Liu X, Fu Y, Liu Z, Lin B, Xie Y, Liu Y, Xu Y, Lin J, Fan X, Dong M, *et al.* (2006) *Am J Hum Genet* 78:393–400.
50. Bamshad MJ, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding S, Stone AC, Jorde LB, Weiss RB, *et al.* (2002) *Proc Natl Acad Sci USA* 99:10539–10544.
51. Hudson RR, Kaplan NL (1988) *Genetics* 120:831–840.
52. Takahata N, Satta Y (1998) *Immunogenetics* 47:430–441.
53. Satta Y, Li Y-J, Takahata N (1998) *Front Biosci* 3:459–467.
54. Grimsley C, Mather KA, Ober C (1998) *Mol Biol Evol* 15:1581–1588.
55. Kelly JK, Wade MJ (2000) *J Theor Biol* 204:83–101.
56. Raymond CK, Kas A, Paddock M, Qiu RL, Zhou Y, Subramanian S, Chang J, Palmieri A, Haugen E, Kaul R, *et al.* (2005) *Gemome Res* 15:1250–1257.
57. Cereb N, Hughes AL, Yang SY (1997) *Immunogenetics* 47:30–36.
58. Fu YG, Liu ZH, Lin JH, Chen WM, Jia ZJ, Pan DJ, Xu AL (2003) *Immunogenetics* 54:761–766.
59. Reusch TBH, Langefors A (2005) *J Mol Evol* 61:531–541.
60. Garrigan D, Edwards SV (1999) *Mol Biol Evol* 16:1599–1606.
61. Cowell LG, Kepler TB, Janitz M, Lauster R, Mitchison NA (1998) *Gemome Res* 8:124–134.
62. Beaty JS, Sukiennicki T, Nepom GT (1999) *Microbes Infect* 1:919–927.
63. Haas JP, Kimura A, Andreas A, Hochberger M, Keller E, Brunnler G, de la Paz Bettinotti M, Nevinny-Stickel C, Hildebrandt B, Sierp G, *et al.* (1994) *Hum Immunol* 39:31–40.
64. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) *Nucleic Acids Res* 31:3497–3500.
65. Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003) *Bioinformatics* 19:2496–2497.
66. Posada D, Crandall KA (1998) *Bioinformatics* 14:817–818.
67. Huelsenbeck JP, Ronquist F (2001) *Bioinformatics* 17:754–755.
68. Page RDM (1994) *Syst Biol* 43:58–77.
69. Page RDM (1998) *Bioinformatics* 14:819–820.

Fig. 4. Alignment of 47 unique *DQA1* proximal promoter haplotypes from
nine primate species. Sequence names are comprised of the first two
letters of the genus and species names. Numbering above sequence
corresponds to location upstream of translation start site (8-bp
insertion at -22 in Pacy 02 was not included in numbering). Periods
indicate identity with Hosa 01. Dashes indicate insertion/deletion
polymorphism. Known transcription factor binding sites are shaded.

```
        -282          -272          -262          -252          -242
         |             |             |             |             |
Hosa 01  GATTCCAATT GAGTGTCCTC CCTTTATTCA TAGAACAATT CCTCAAGTCC
Hosa 02  .......... T.A....... .......... .......... ..........
Hosa 03  .......... T......... .......... .......... ..........
Hosa 04  .......... TC........ .......... .......... ..........
Hosa 05  ........C. .......... .......... .......... ..........
Hosa 06  .......... T......T.. ...C.C.... .......... ..........
Hosa 07  .......... T......... ...C.G.... .......... ..........
Hosa 08  .......... T.......G ...C.G.... .......... ..........
Hosa 09  ......G... TC........ ...C...... C......T.. ..........
Hosa 10  .......... T......... ...C...... .....T.T.. ..........
Patr 01  .......... T......... ...C...... .....T.T.. ..........
Patr 02  .......... T......... ...C...... .......... ..........
Patr 03  .......... T......... ...C.G.... .......... ..........
Patr 04  .......... T......... ...C...... .......... ..........
Patr 05  .......... T......... ...C.G.... .......... ..........
Papa 01  .......... T......... ...C...... .....T.T.. ..........
Gogo 01  .......... T......... ...C...... .....T.T.. ..........
Gogo 02  ....T..... T......... ...C.G.... .......... ..........
Poab 01  ......G... TC........ ...C...... .......... ..........
Poab 02  ......G... TC........ ...C...... .......... ..........
Poab 03  ......G... TC........ ...C...... .......T.. A.........
Poab 04  .......... TC........ ...C...... .......T.. A.........
Poab 05  .......... T......... ...C...... .......T.. ..........
Popy 01  .......... TCA....... ...C...... .......T.. ..........
Popy 02  ......G... TCA....... ...C...... .......T.. ..........
Hysy 01  ...C...G.. T......... ...C..G... ...C...T.. ..........
Hysy 02  ...C...G.. T......... ...C..G... ...C...T.. ..........
Hysy 03  ...C...G.. T......... ...C..G... ...C...T.. ..........
Hysy 04  .......... TC........ ...C...... C..C...T.. ..........
Pacy 01  .......... T....C.... ...C...... .......T.. ..........
Pacy 02  .......... T......... ...C...... .......T.. ........G.
Pacy 03  .......... T......T.. ...C...... .......T.. .........T
Pacy 04  .......... T......... ...C...... .......T.. ..........
Pacy 05  .......... T......... ...C...... .......T.. ..........
Pacy 06  .......... T....C.... ...C...... .......T.. ..........
Pacy 07  .......... T.....T.T ...C...... .......T.. .........T.
Pacy 09  .......... T......... ...C...... .......T.. ..........
Pacy 10  .......... T......... ...C...... .......T.. ..........
Pacy 11  .......... T......... ...C...... .......T.. ..........
Pacy 12  .......... T......... ...C...... .......T.. ..........
Pacy 13  .......... T......... ...C...... .......T.. ..........
Mamu 01  T......... T.....T.. ...CC..... ...G..T.. .........T
Mamu 02  .......... T....T... ...C...... .......T.. ..........
Mamu 03  .......... T....T... ...C...... .....T.T.. ..........
Mamu 04  .......... T......... ...C...... .......T.. ..........
Mamu 05  .......... T......... ...C...... C......T.. ..........
Mane 01  .......... T....T... ...C...... .......T.. ..........
```

```
          -232         -222         -212         -202         -192
           |            |      W      |            |  NF-_B     |
Hosa 01   ACTCTGAGTA GAGGCTGCAT CACAA---GG GGATTGCCCC GTCTCCTTCC
Hosa 02   .......... .......... .....---.. .......... ..........
Hosa 03   .......... .......... .....---.. .......... ..........
Hosa 04   .......... .......... .....---.. .......... ..........
Hosa 05   .......... .......... .....---.. .......... ..........
Hosa 06   .......... .......... .....CAA.. .........T ..........
Hosa 07   .......... .......... .....CAA.. .........T ..........
Hosa 08   .......... .......... .....CAA.. .........T ..........
Hosa 09   ........G. .......... .....CAA.. .........T ..........
Hosa 10   ........G. .......G. .....CCA.. .....T...T ..........
Patr 01   ........G. .......G. .....CCA.. .....T...T ..........
Patr 02   .......... .......... .....CAA.. .......... A.........
Patr 03   .......... .......... .....CAA.. .........T ..........
Patr 04   ........G. .......G. .....CCA.. .....T...T ..........
Patr 05   .......... .......... .....CAA.. .........T ..........
Papa 01   ........G. .......G. .....CCA.. .....T...T ..........
Gogo 01   ........G. .......G. .....CCA.. .....T...T ..........
Gogo 02   ........G. .......G. .....CCA.. .....T...T ..........
Poab 01   .......... .......... .....CCA.. .........T .........T.
Poab 02   .......... .......... .....CAA.. .........T ..........
Poab 03   ........G. .......... .....CCA.. .....T...T ..........
Poab 04   ........G. .......... .....CCA.. .....T...T ..........
Poab 05   ........G. .......... .....CCA.. .........T .........T.
Popy 01   ........G. .......... .....CCA.. .....T...T ..........
Popy 02   ........G. .......... .....CCA.. .....T...T ..........
Hysy 01   .......CG. .......... .....CCA.. .........T .....T....
Hysy 02   .......CG. .......... .....CCA.. .....T...T .....T....
Hysy 03   .......CG. .......... .....CCA.. .....T...T .....T....
Hysy 04   .......C.. .......... .....CCA.. .........T ..........
Pacy 01   ........G. .......... .....CCA.. .........T ..........
Pacy 02   ........G. .......... .....CCA.. .........T A....TCC..
Pacy 03   ........G C......... .....CCA.. .........T ..........
Pacy 04   ........G. ...A...... .....CCA.. .........T ..........
Pacy 05   .......A.. ..C....... .....CCA.. .........T .......C..
Pacy 06   ........G. .......... .....CCA.. .........T ..........
Pacy 07   ........G. .......... .....CCA.. .........T ..........
Pacy 09   .......A.. .......... .....CCA.. .........T .......C..
Pacy 10   ......A.. .......... .....CCA.. .........T .......C..
Pacy 11   ........G. .......... .....CCA.. .........T ..........
Pacy 12   .......A.G .......... .....CCA.. .........T .......C..
Pacy 13   .......A.. .......... .....CCA.. .........T .......C..
Mamu 01   ........G. ...A...... .....CCA.. .........T ..........
Mamu 02   .......A.. .......... .....CCA.. .........T .......C..
Mamu 03   ......A.. .......... .....CCA.. .........T .......C..
Mamu 04   ......A.. .......... .....CCA.. .........T .......C..
Mamu 05   .......A.G .......... .....CCA.. .........T .......C..
Mane 01   .......A.. .......... .....CCA.. .........T .......C..
```

```
          -182        -172        -162        -152        -142
          |  S        |           |  J        |  X1       |X2
Hosa 01   AGGGCTCTTA ATACAAACTC TTCAACTAGT AACTGAGATG TCACCATAGG
Hosa 02   .......... .......... .......... .......... ..........
Hosa 03   .......... .......... .......... .......... ..........
Hosa 04   .......... .......... .......... .......... ..........
Hosa 05   .......... .......... .......... .......... ..........
Hosa 06   .......... ...G...... .......... .......... .......G...
Hosa 07   .......... .......... ....G..... .......... .......G...
Hosa 08   .......... .......... .......... .......... .......G...
Hosa 09   .......... .......... .......... ......G... ...T......
Hosa 10   .......... .......... .......... ......G... ...T......
Patr 01   .......... .......... .......... ......CG.. ...T......
Patr 02   .......... .......... .......... .......... ..........
Patr 03   .......... .......... .......... ......A... .......G..
Patr 04   .......... .......... .......... ......CG.. ...T......
Patr 05   .......... .......... .......... ......A... .......G..
Papa 01   .......... .......... .......... ......G... ...T......
Gogo 01   .......... .......... .......... ......G... ...T......
Gogo 02   .......... .......... .......... ......G... ...T......
Poab 01   .......... .......... .......... ......G... ....T.....
Poab 02   .......... .......... .......... .......... ..........
Poab 03   .......... .......... .......... ......G... ...T......
Poab 04   .......... .......... .......... ......G... ...T......
Poab 05   .......... .......... .......... ......G... ....T.....
Popy 01   .......... ......T.. .......... ......G... ...T.....A
Popy 02   .......... .......... .......... ......G... ...T......
Hysy 01   .......... .......... .......... .....AG.. ...T......
Hysy 02   .......... .......... .......... .....AG.. ...T......
Hysy 03   ......T... .......... .......... .....AG.. ...T......
Hysy 04   .......... .......... .......... ......G... ....T.....
Pacy 01   .......... .......... .......... ......G... ..........
Pacy 02   G......... .......... ..A....... ......G... ..........
Pacy 03   .......... .......... .......... ......G... ..........
Pacy 04   .......... .......... .......... ......G... ..........
Pacy 05   .......... .......... .......... ..T....G.. ...T..G...
Pacy 06   .......... ......T.. .......... ......G... ..........
Pacy 07   ......T... ..T....... .......... ......G... ..........
Pacy 09   .......... .......... .......... ..T....G.. ...T..G...
Pacy 10   .......... .......... .......... ..T....G.. ...T..G...
Pacy 11   .......... .......... .......... ......G... ...T......
Pacy 12   .......... .......... .......... ..T...G.. ...T..G...
Pacy 13   .......... .......... .......... ..T...G.. ...T..G...
Mamu 01   .......... .......... .......... ......G... ..........
Mamu 02   .......... .......... .......... ..T....G.. ...T..G...
Mamu 03   .......... .......... .......... ..T....G.. ...T..G...
Mamu 04   .......... .......... .......... ..T...AG.. ...T......
Mamu 05   .......... .......... ...C...... ..T....G.. ...T......
Mane 01   .......... .......... .......... ..T....G.. ...T..G...
```

```
          -132          -122          -112          -102          -92
           |            |  Y          |            |            |
Hosa 01   GGATTTTTCT AATTGGCCAA AACCTGACTT GGCAGGGTTT GGTTTGGGTG
Hosa 02   .......... .......... .......... .......... ..........
Hosa 03   .......... .......... .......... .......... ..........
Hosa 04   .......... .......... .......... .......... ..........
Hosa 05   .......... .......... .......... .......... ..........
Hosa 06   .......... .......... .......... .......... ..........
Hosa 07   .......... .......... ........C. .......... ..........
Hosa 08   .......... .......... ........C. .......... ..........
Hosa 09   .......... ....A..... .......... .......... ..........
Hosa 10   ...G...... ....A..... ........C. .......... ..........
Patr 01   ...G...... ....A..... ........C. .......... ..........
Patr 02   .......... .......... ........C. .......... ..........
Patr 03   .......... .......... ........C. .......... ..........
Patr 04   ...G...... ....A..... ........C. .......... ..........
Patr 05   .......... .......... ........C. .......... ..........
Papa 01   ...G...... ....A..... ........C. .......... ..........
Gogo 01   ...G...... ....A..... ........C. .......... .........A
Gogo 02   ...G...... ....A..... ........C. .......... .........A
Poab 01   .......... .......T... ........C. ....A..... ..........
Poab 02   .......... .......... ........C. .....A.... .....C.....
Poab 03   .......... ....A..... ........C. .......... .A........
Poab 04   .......... ....A..... ........C. ..T....... ..........
Poab 05   .......... .......T... ........C. ....A..... ..........
Popy 01   .......... ....A..... ........C. .......... ..........
Popy 02   ........T. ....A..... ........C. .......... ..........
Hysy 01   .T........ .......... ........C. .......... ..........
Hysy 02   .......... .......... ........C. ....A..... ..........
Hysy 03   .......... .......... ........C. ....A..... ..........
Hysy 04   .......... .......T... ........C. .......... .A........
Pacy 01   .......... .......T... .......TC. ......T..G ........C.
Pacy 02   .......... .......T... .......TC. ....A..... ..........
Pacy 03   .......... .......T... .......T.. .......... ..........
Pacy 04   .......... .......T... .......TC. ......T..G ........C.
Pacy 05   A......... G......... ...T....C. .......... .........A
Pacy 06   .......... .......T... .......TC. ......T..G ........C.
Pacy 07   .......... .......... .......AC. .......... ..........
Pacy 09   A......... G......... ...T....C. .......... .........A
Pacy 10   A......... G......... ...T....C. .......... .........A
Pacy 11   .......... ..C....... ...A....C. ...G...... .....A....
Pacy 12   A......... G......... ...T....C. .......... .........A
Pacy 13   A......... G......... ...T....C. .......... .........A
Mamu 01   .......... .......T... .......T.. .......... ..........
Mamu 02   A......... G......... ...T....C. .A......C. .........A
Mamu 03   A......... G......... ...T....C. .......... .........A
Mamu 04   A......... G......... ...T....C. .A......C. .........A
Mamu 05   A......... G......... ...T....C. .A......C. .........A
Mane 01   A......... G......... ...T....C. .......... .........A
```

4

```
              -82        -72        -62        -52        -42
               |          |          |          |          |
Hosa 01   TCTTCAGATT TCCTTGTCTT GAGGTCCTCA CAATTACTCT ACAGCTCAGA
Hosa 02   .......... .......... .......... .......... ..........
Hosa 03   .......... .......... .......... .......... ..........
Hosa 04   .......... .......... .......... .......... ..........
Hosa 05   .......... .......... .......... .......... ..........
Hosa 06   .......... .......... .......... .......... ..........
Hosa 07   .......... .......... ....C..... .....G.... ..........
Hosa 08   .......... .......... ....C..... .....G.... ..........
Hosa 09   .......... G.......C .......... .....G.... ...A......
Hosa 10   .......... .G......C .......... .....G.... ..........
Patr 01   .......... ...C....C ...A...... .....G.... ...A......
Patr 02   .......... .......... .......... .......... ....T.....
Patr 03   .......... .......... ........T. .......... ..........
Patr 04   .......... ...C....C ...A...... .....G.... ...A......
Patr 05   .......... .......... ........T. .......... ..........
Papa 01   .......... ...C....C ...A...... .....G.... ...A......
Gogo 01   .......... ........C ...A...... .....G.... ...A......
Gogo 02   .......... ........C ...A...... .....G.... ...A......
Poab 01   .......... ........C .......... .....G.... ..........
Poab 02   .......... .......... .......... .......... ..........
Poab 03   .......... ........C .......... ...A.G.... ...A......
Poab 04   .......... ........C ...A...... .....G.... ...A......
Poab 05   .......... ........C .......... .....G.... ..........
Popy 01   .......... A.......C A.C....... .....GA... ...A......
Popy 02   .......... G.......C A......... .....G.... ...A......
Hysy 01   .......... ........C A......... .....G.... ...A......
Hysy 02   .......... ........C .......G.. .....G.... ...A......
Hysy 03   .......... ........C .......G.. .....G.... ...A......
Hysy 04   .......... ......C..C A......... ..G..G.... ...A......
Pacy 01   .......... ........C .......... .......... ..........
Pacy 02   .......... .T......C A......... .......... ...A......
Pacy 03   .......... .......... .......... .......... ..........
Pacy 04   .......... ........C .......... ........G. ..........
Pacy 05   ....T..... ........C .......... .....G.... .G........
Pacy 06   .......... ........C .......... .......... ..........
Pacy 07   .......... ........C C.A....... .......... ..........
Pacy 09   .......... ........C .......... ...C.G.... .G........
Pacy 10   ....T..... ........C .......... .....G.... .G........
Pacy 11   .......... ........C .......... .....G.... ..........
Pacy 12   ....T..... ........C .......... T....G.... .G........
Pacy 13   ....T..... ........C .......... .....G.... .G........
Mamu 01   .......... .......... .......... .......... ..........
Mamu 02   ....T..... ........C .......... .....G.... .G........
Mamu 03   .......... ........C .......... .....G.... .G........
Mamu 04   ....T..... ........C .......... .....G.... .G........
Mamu 05   ....T..... ........C ........T. .....G.... .G........
Mane 01   .......... ........C .......... .....G.... .G........
```

5

```
                 -32                -22       -13
                  |                  |         |
Hosa 01    ACACCAACTG --------CT GAGGCTGC
Hosa 02    .......... --------.. ........
Hosa 03    .......... --------.. ........
Hosa 04    .......... --------.. ........
Hosa 05    .......... --------.. ........
Hosa 06    ...G...... --------.. ........
Hosa 07    ...G...... --------.. ........
Hosa 08    ...G...... --------.. ........
Hosa 09    ...G...... --------.. A.......
Hosa 10    G..G...... --------.. ........
Patr 01    G..G...... --------.. ........
Patr 02    .......... --------.. ........
Patr 03    ...G...... --------.. ........
Patr 04    G..G...... --------.. ........
Patr 05    ...G...... -------.. .......T
Papa 01    G..G...... --------.. ........
Gogo 01    G..G...... --------.. A.......
Gogo 02    G..G...... --------.. A.......
Poab 01    ...G...... -------T. ........
Poab 02    ...G...... --------.. ........
Poab 03    G..G...... --------.. A.......
Poab 04    G..G...... --------.. A.......
Poab 05    ...G...... -------T. ........
Popy 01    .T.G...... --------.. A.......
Popy 02    ...G...... --------.. A.......
Hysy 01    ...G...... --------.. ........
Hysy 02    .T.G...... --------.. ........
Hysy 03    .T.G...... --------.. ........
Hysy 04    ...G...... --------.. ........
Pacy 01    ...G...... --------.. ........
Pacy 02    ...G...... AGCAACTG.. ........
Pacy 03    ...GT..... --------.. ........
Pacy 04    ...G...... --------.. ........
Pacy 05    ..GG...... --------.. ..A...A.
Pacy 06    ...G...... --------.. ........
Pacy 07    ...G..G..T --------.. ........
Pacy 09    ..GG...... --------.. ........
Pacy 10    ..GG...... --------.. ..A...A.
Pacy 11    ...G...T.. --------.. ........
Pacy 12    ..GG...... --------.. ......A.
Pacy 13    ..GG...... --------.. ......A.
Mamu 01    ...G...... --------.. ........
Mamu 02    ..GG...... --------.. ........
Mamu 03    ..GG...... --------.. ........
Mamu 04    ..GG...T.. --------.. ........
Mamu 05    ..GG...... --------.. ........
Mane 01    ..GG...... --------.. ........
```