Title

Natural Selection on HLA-DPB1 Amino Acids Operates Primarily on DP Serologic Categories

Journal

Human Immunology, 85(6)

ISSN

0198-8859

Authors

Single, Richard M
Mack, Steven J
Solberg, Owen D
et al.

Publication Date

2024-11-01

DOI

10.1016/j.humimm.2024.111153

Title:
Natural Selection on HLA-DPB1 Amino Acids Operates Primarily on DP Serologic Categories

Authors:
Richard M. Single[1*], Steven J. Mack[2*], Owen D. Solberg[3], Glenys Thomson[4], Henry A. Erlich[5]

Author Affiliations:
1 Department of Mathematics and Statistics, University of Vermont, Burlington, VT
2 Department of Pediatrics, University of California, San Francisco, Oakland, CA
3 Bioinformatics and Biostatistics, Monogram Biosciences, South San Francisco, CA
4 Department of Integrative Biology, University of California, Berkeley, CA
5 Center for Genetics, Children's Hospital & Research Center Oakland, Oakland, CA

* These authors contributed equally to the work described.

Corresponding Author:
Steven J. Mack
Department of Pediatrics
University of California, San Francisco
5700 Martin Luther King Jr. Way
Oakland, CA 94609
Phone: 510-597-7145
Fax: 510-450-7910
steven.mack@ucsf.edu

Abbreviated Title: Selection on DP Serologic Categories

Abstract

The *DPB1* locus is notable among the classical *HLA* loci in that allele frequencies at this locus are consistent with genetic drift, whereas the frequencies of specific DPβ amino acids are consistent with the action of balancing selection. We investigated the influence of natural selection in shaping the diversity of three functional categories of *DPB1* diversity defined by specific amino acid motifs, *DPB1* T-cell epitopes, *DPB1* supertypes and DP1-DP4 serologic categories (SCs), via Ewens-Watterson (EW) selective neutrality and asymmetric Linkage Disequilibrium (ALD) analyses in a worldwide sample of 136 populations. These EW analyses provide strong evidence for the operation of balancing selection on DP SCs, but no evidence for balancing selection on T-cell epitopes or supertypes. We further investigated the global distribution of SCs. Each SC is common in a different region of the world, with the DP1 SC most common in Southeast Asia and Oceania, the DP2 SC in North and South America, the DP3 SC in South America, and the DP4 SC in Europe. The DP2 SC is present in all populations, while 14% of populations are missing at least one DP1, DP3, or DP4 SC. We observed consistent *DPA1*~DP SC haplotype associations across 10 populations from five global regions, and found that asymmetric linkage disequilibrium (LD) between the *DPB1* locus and the four most-common *DPA1* alleles (*DPA1*01:03*, **02:01*, **02:02* and **03:01*) is determined by variation at DPβ AA positions 85-87. These positions are in LD with both DPα positions 31 and 50. We conclude from these EW analyses that natural selection is primarily operating to maintain population-level diversity of DP SCs, rather than *DPB1* alleles or other functional categories of *DPB1* diversity.

Keywords:
DPB1; Balancing Selection; DP Serologic Categories; Amino Acid; Population Study

Abbreviations:
AA: Amino Acid
af: allele frequency
AFND: Allele Frequency Net Database
ALD: asymmetric linkage disequilibrium
AUS: Australia
EUR: Europe
EW: Ewens-Watterson
GD: Genotype Dataset
GMT: Generic Mapping Tools
LD: Linkage Disequilibrium
NAF: North Africa
NAM: North America
NEA: Northeast Asia
OCE: Oceania
OTH: Other
RT: Randomization Test
SAM: South America
SC: Serologic Category
SEA: Southeast Asia
SLDC: Solberg Literature Dataset Compilation
SSA: Sub-Saharan African
ST: Supertype
SWA: Southwest Asia
TCE: T-Cell Epitope

1. Introduction

The HLA, so-called “human leukocyte antigen”, cell-surface proteins play a role in distinguishing self from non-self peptides by presenting intra- and extracellular-derived peptides to T-cell receptors. Located on chromosome 6p21.3, the classical class I and II genes are the most polymorphic loci in the human genome. Extensive linkage disequilibrium (LD) is found both within and between the HLA class I and class II gene regions[1-3]. Specific HLA alleles, allele-families and haplotypes have been associated with susceptibility to and protection from pathogens, auto-immune diseases, and cancers [4-11].

The allelic diversity of the HLA loci has been shaped by natural selection [12]. Allele frequency distributions at the *HLA-A*, *-C*, *-B*, *-DRB1*, *-DQA1*, and *-DQB1* loci are generally more even than expected under neutral conditions, a pattern consistent with balancing selection [1, 2, 13-22], while those at the *DPB1* locus are generally compatible with neutral evolution via genetic drift, with evidence for directional selection in a few cases [14, 15, 19-21, 23, 24]. Insufficient population data are available to draw conclusions regarding the strength of selection at the DPA1 locus, although Solberg et al. [20] suggested that balancing selection at this locus was between *HLA-A* and *–B* in strength.

In the companion paper, we applied the Ewens-Watterson (EW) homozygosity test of neutrality to frequency distributions of *DPB1* alleles and polymorphic *DPB1* exon 2-encoded amino acid (AA) positions, as well as to pairs and trios of polymorphic AA positions, based on averages over populations. We found that 64% of polymorphic *DPB1* AA positions (8, 9, 11, 36, 55, 56, 69, 84 and 85+, the last representing a dimorphic trio of either a E85-A86-V87 or a G85-P86-M87 motif) were evolving under balancing selection [companion paper]. Site-directed mutagenesis experiments have revealed AA positions 9, 11, 36, 55, 56, 69, 84, 85, 86 and 87 to be central to the functions of the DP molecule [25, 26]. In the companion paper, we proposed that the failure of EW analysis to detect balancing selection for the DPB1 locus may indicate that, unlike alleles of other HLA loci, DPB1 alleles do not represent the functional categories of DPB1 diversity. Here, we apply EW analyses at the individual population level to investigate the action of natural selection on the three functionally-defined categories of DPB1 diversity – T-cell epitopes (TCEs), supertypes (STs) and serologic categories (SCs). TCEs can be used to categorize HLA alleles based on their ability to present peptides from antigens recognized by T-cells. STs are groups of HLA alleles that share similar structural features and, thus, similar peptide-binding specificities. SCs classify alleles based on the reaction of HLA molecules with specific antibodies.

Based on their recognition by alloreactive T-cell clones, *DPB1* alleles can be subdivided into three (TCE3) or four (TCE4) TCE groups. TCEs encoded by distinct sub-sets of DPB1 alleles have proven relevant for unrelated hematopoietic stem-cell transplantation [27-30]. Defined by polymorphisms at DPβ AA positions 11, 69, and 84 that impact peptide-binding, six DP STs (DP1, DP2, DP3, DP4, DP6, and DP8), have been implicated in susceptibility to childhood acute lymphoblastic leukemia [31, 32]. Dimorphic AA variants at DPβ AA positions 56 and 85+ have been identified as the primary immunodominant

serologic epitopes of the DP molecule, allowing all *DPB1* alleles can be divided into four SCs (DP1, DP2, DP3, and DP4) [33].

The action of balancing section acting on these DP SCs has been inferred by Voorter et al. in the population of Guadeloupe[34], and by Hollenbach et al. in a European American population of 6000 individuals[35]. Hollenbach et al. further inferred selection for stable DP heterodimer formation to be determined by interactions between DPα AA 31 and DPβ AAs 85-87, which are proximal in the DP2 crystal structure, based on LD between these AA positions.

We compare patterns of natural selection inferred for the functionally-defined TCEs, STs and SCs to *DPB1* alleles, their encoded AA positions and AA motifs in a set of 136 population samples, representing a world-wide sample of 13,338 individuals. Finally, we present an overview of the distribution and prevalence of DP SCs in this sample.

2. Materials and Methods

2.1. Population samples

The analyzed dataset represents a global sampling of 13,338 individuals from 136 populations originally published in anthropological studies or as healthy control populations for case-control studies [2, 16, 23, 36-99]. Each individual population dataset has been subjected to quality control scrutiny, and the overall dataset reviewed to eliminate duplications [20]. The three primary sources for these data are described in the companion paper -- *Solberg Literature Dataset Compilation (SLDC), Allele Frequencies Net Database (AFND),* and *Genotype Datasets (GD).*

DPA1 genotype data for 10 of these populations were obtained from the former dbMHC database (ftp.ncbi.nlm.nih.gov/pub/mhc/mhc/Final%20Archive/IHWG/Anthropology), and represent 739 individuals from sub-Saharan Africa, Europe, Oceania, North America, and South America.

2.2. Data Analysis

2.2.1. Software

The Python for Population Genomics (PyPop, version 0.7.0, www.pypop.org) [100, 101], the R statistical environment (version 3.0.1) and the asymLD R package (v0.1, https://cran.r-project.org/web/packages/asymLD)[102, 103] are described in the companion paper.

Maps of interpolated DP SC frequency distributions were generated using the Generic Mapping Tools (GMT) package [104] (version 4) via the blockmean and surface functions as described by Solberg et al. [20].

The HLA-DP2 (*DPA1*01:03*, *DPB1*02:01*) protein crystal structure [105] (Protein Data Bank ID 3LQZ), was obtained from the National Center for Biotechnology Information's Molecular Modeling Database (http://www.ncbi.nlm.nih.gov/structure?term=DPB1), and manipulated in CN3D v4.3.1 [106].

2.2.2. Standardization of DPA1 and DPB1 alleles across population datasets

The *DPB1* allele names and sequences in Immuno Polymorphism Database (IPD)-ImMunoGeneTics (IMGT)/HLA database (version 3.4.0) were used for all comparisons and analyses. *DPB1* allele names were validated and translated to version 3.4.0 names using the Allele Name Translation Tool (version 0.5.0) [107]. *DPA1* and *DPB1* alleles with identical exon 2 nucleotide sequences were combined into a common allele category for analysis. Allele names that included more than two polymorphic fields (e.g. *DPB1*01:01:01*) were truncated to two fields (e.g. *DPB1*01:01*); all *DPA1* and *DPB1* allele-level analyses were carried out at the protein-level. The same rules for consistent nomenclature, data validation, and ambiguity resolution were applied to datasets from each of the three sources. These rules are available in the config-allelecount.ini configuration file available at http://pypop.org/popdata/.

2.2.3. Definition of locus-categories

Based on the AA sequences for each allele name reported in the dataset, *DPB1* alleles were assigned to the following distinct "locus-categories" for analysis: TCE3 and TCE4 TCE groups, DP STs, and DP SCs. This process, referred to as "collapsing" *DPB1* alleles to a specific locus-category, is described below.

DPB1 alleles were assigned to TCE3 and TCE4 groups as detailed in Table1. The TCE3 and TCE4 group status for the alleles in the dataset analyzed here is provided in Table 2. Although Zino et al. [28] and Sizzano et al. [27] differ in the assignment of *DPB1*86:01* and **104:01* to TCE3 groups 2 and 3, homozygosity calculations for both TCE3 group definitions differed only in the Tunisia population, and insignificantly (F = 0.5826 vs. 0.5827). We performed and report on analyses for TCE3 groups as defined by Sizzano et al. [27] here.

DPB1 alleles were assigned to DP ST categories as described in Tables 1 and 2. When these six ST categories were originally defined [31] AA positions 11, 69, and 84 were treated as dimorphic, and though only six DP STs are defined, up to eight categories are possible for three sets of dimorphisms. In addition, three residues (E, K and R) have been observed for position 69, whereas the ST definitions involve only E69 and K69, and three residues (D, G, and V) have also been observed for position 84, whereas ST definitions involve only G84 and D84. Overall, 18 possible residue triplets are possible for this trio. Of the alleles in this dataset, R69 is present in *DPB1*11:01*, **15:01* and **69:01*, and V84 is present in *DPB1*15:01, *18:01, *28:01, *34:01, *40:01 and *62:01*. Therefore, comparisons for these AA positions include four observed residue triplets (L11:K69:G84, G11:K69:V84, G11:R69:V84, and L11:R69:D84) in addition to the residue triplets corresponding to the six DP STs, as shown in Table 1.

DPB1 alleles were assigned to the DP SCs as described in Tables 1 and 2.

2.2.4. Analytical Methods

The tests of neutrality, haplotype frequency estimation, and linkage disequilibrium statistics used in this study are described in the companion paper. Details specific to analyses of categories, types, and groups described in this paper are presented below.

For the analyses of each locus-category (DP TCEs, DP STs and DP SCs), each variant in a given locus-category was treated as a discrete allele-category. For example, in the analysis of the SCs, all *DPB1* alleles encoding A56, E85, A86 and V87 were collapsed into the DP1 SC, while all alleles encoding E56, G85, P86 and M87 were collapsed into the DP2 SC. EW homozygosity test (EW test) of neutrality statistics (F_{obs}, F_{exp}, and F_{nd}) were computed based on the frequencies for the allele-categories as described in the companion paper. We used the t.test function in the R *Stats* package to perform two-tailed t-tests to determine if mean F_{nd} values differed significantly from the null-hypothesis of neutrality (F_{nd} = 0). We refer to this below as a "neutrality t-test".

The low number (4) of DP SCs relative to *DPB1* alleles for a given population makes direct comparison of p-values problematic, as the EW test is affected by the number of alleles at a locus. In addition to the EW test of significance described above, we developed a resampling approach that controls for the restricted number of allele-categories for the DP SCs in each population (DP SC resampling). The observed F_{nd} values for DP SCs in each population were compared to a set of 5,000 randomized F_{nd} values that were calculated for each population after randomly assigning each observed allele in that population to one of four categories, in the same proportions with which alleles were assigned to the true DP serologic categories, controlling for the number of alleles/categories. A randomization p-value was computed for each population sample as the proportion of randomized F_{nd} values that were lower than the true observed F_{nd} value for the DP SCs. We performed similar resampling analyzes for DP ST and TCE3 groups.

We used PyPop's expectation-maximization (EM) algorithm to estimate *DPA1~DPB1* haplotype frequencies and *DPA1*~DP SC haplotype frequencies on the basis of *DPA1* and *DPB1* genotype data, and inferred DP SCs in 10 populations using 50 starting conditions. In addition, we compared *DPA1*~DP SC haplotype frequencies estimated via the EM algorithm (estimated *DPA1*~DP SC haplotypes) to the frequencies of DPA1~DP SC haplotypes inferred from *DPA1~DPB1* haplotypes (inferred *DPA1*~DP SC haplotypes) by collapsing the *DPB1* allele in each *DPA1~DPB1* haplotype to a DP SC as described in section 2.2.3.

We calculated the normalized allele-level LD measure ($D'_{ij} = D_{ij}/D_{max}$ [108]) in PyPop, where D_{max} is the maximum value that D_{ij} can achieve. D'_{ij} ranges from -1 to +1, with a D'_{ij} value of 0 indicating linkage equilibrium. A value of +1 may indicate the complete association of a given pair of alleles in a single haplotype, and a value of -1 may indicate the complete absence of a haplotype comprised by those alleles.

We calculated the conditional asymmetric LD (*ALD*) statistics, $W_{A|B}$ and $W_{B|A}$ [102, 103], which complement the global LD measure, $W_{n\ [109]}$ (not reported), in cases when loci display different numbers of alleles as described in the companion paper.

2.2.4.4. Correction for Multiple Comparisons

In the tables we report uncorrected p-values, with p-value threshold for a Bonferroni correction based on the number of tests performed listed in each table. This p-value

threshold is included as a conservative reference value, and represents an overcorrection, as these tests are not independent due to correlations from LD and shared population histories.

3. Results

3.1. Alleles Surveyed

The compiled dataset included 8 *DPA1* alleles (DPA1*01:03, 01:04, 02:01, 02:02, 02:03, 03:01, 03:02, 04:01) and the 74 *DPB1* alleles shown in Table 2. Of the 74 observed *DPB1* alleles, only the *DPB1*61:01N* allele could not be collapsed into a locus-category for analysis. This allele was reported once in the Cameroon dataset, and in no other populations. The frequency of each *DPA1* and *DPB1* allele in each population is presented in Supplementary Table S1.

3.2. Amino acid-Level Analyses of Selection

As summarized in the companion paper, F_{nd} values were calculated for each polymorphic AA position, each pair of AA positions, and each trio of AA positions across all populations. Detailed results are presented in Table 3 and Supplementary Table S2, Figures 4 and 5 and Supplementary Table S3, respectively.

These patterns of amino acid level F_{nd} variation are consistent with those reported by Salamon et al. [14] in 14 populations, and by Lancaster [110] in 22 populations (all analyzed here), where low F_{nd} values were observed in three distinct regions of the AA sequence. Eighteen AA positions in the compiled dataset were polymorphic, and four (12, 17, 32, and 72) were monomorphic in most populations and excluded from subsequent analyses. Position 33 was polymorphic in only 51% of populations. No population displayed a significantly low p-value for these five positions, and subsequent AA analyses pertain to the remaining 13 positions.

3.3 Analyses of Selection on Functional Categories

Summaries of the analyses of selection are presented in Table 4, and are described below. Individual F_{nd} values and their associated p-values for each population and each functional category are provided in Supplementary Table S5, alongside the corresponding value for *DPB1* alleles in each population (originally presented in the companion paper). *DPB1* allele-frequency distributions have been previously shown to be consistent with the null hypothesis of neutral evolution ($F_{nd} = 0$) observed [14, 15, 20][companion paper], and values for functional categories of *DPB1* polymorphism should be considered in that context.

3.3.1 T-Cell Epitopes

The mean F_{nd} values for TCE3 and TCE4 groups were 0.17 and -0.26, respectively. Individual F_{nd} values were only significant for the Coreguaje and Bari populations (p-values = 0.0202 and 0.0363, respectively for both TCE3 and TCE4 groups), both of which were missing TCE3 group 1 and TCE4 groups 1 and 3 alleles. TCE3 group 1 and 2 alleles were absent from five populations (Kimberly, Mixe, Pima_17, Tolai_1999, and TrobriandIslanders_1999), and 35% of populations were missing one TCE3 group. Thirteen percent of populations were missing two TCE4 groups. Of these, all were

missing TCE4 group 1 alleles, 14 of these were also missing TCE4 group 3 alleles, and five were missing TCE4 group 2 alleles. Twenty-six percent of populations were missing one TCE4 group. When compared via the neutrality t-test, TCE3 F_{nd} values were nonsignificant after correction (p-value = 0.006) while those for TCE4 were slightly significant (p-value = $9x10^{-5}$) in the direction of negative homozygosity.

3.3.2 DPB1 Supertypes

The mean F_{nd} value for DPB1 STs and the L11:K69:G84, G11:K69:V84, G11:R69:V84, and L11:R69:D84 residue triplets was -0.495. When compared to the mean F_{nd} values of 363 other AA trios (described in the companion paper), this value ranked 205; the 1st ranked F_{nd} value (for the 55-56-57 trio) was -1.062, and the 364th ranked value (for the 9-57-76) trio was 1.003. Significant individual p-values were observed for 18 populations (13%). No population displayed all 10 STs and residue triplets; five populations (AfricanAmerican_1997, BlacksUS_2003, Gabonese_1998, Kenyan_142, Shona, and Zulu) displayed nine triplets, and nine populations (Arsario_1996, Coreguaje_1996, Ijka_1996, Kogui_1996, Pima_17, Vaupes_1996, Warao_2004, Yucpa_2001, and Yucpa_2004) displayed only two. Eighty-six populations displayed at least one of the four non-ST triplets, averaging 1.65 each, with a maximum of three observed in 11 populations (AfricanAmerican_1997, BlacksUS_2003, British_1994b, EcuadorianAfricans_2001, Gabonese_1998, Kenyan_142, Martinique_2001, Shona, Tunisia_1995, Tunisian_2004 and Zulu). On average, 1.044 non-ST triplets and 5.132 STs were observed over all 136 populations. When compared via the neutrality t-test, F_{nd} values trended significantly in the direction of negative homozygosity (p-value = 1.3×10^{-10}).

3.3.3 DP Serologic Categories

The mean F_{nd} value for DP SCs was -1.152, and 73 populations (54%) displayed significant homozygosity values for DP SCs. One-hundred eighteen populations displayed all four DP SCs, while nine displayed three (Embera_1996, Kimberley, Mataco_1992, Mixe, Pumi_2002, Taiwanese_1999, Tolai_1999, TrobriandIslanders_1999and Wayuu_1996) and nine displayed two (Arsario_1996, Coreguaje_1996, Ijka_1996, Kogui_1996, Pima_17, Vaupes_1996, Warao_2004, Yucpa_2001 and Yucpa_2004) DP SCs. One-hundred twenty-four populations displayed negative F_{nd} values, and when compared via the neutrality t-test, F_{nd} values were highly significant in the direction of negative homozygosity (p-value = 2.93×10^{-39}).

Given that each DP SC represents a pair of AA polymorphisms at DPB1 positions 56 and 85+, whereas *DPB1* alleles represent combinations of multiple polymorphic positions, it is possible that the EW analysis of any pair of polymorphic AA positions may yield results similar to those for DP SCs. As discussed in the companion paper, and presented in Supplementary Table S2, AA position pair 36:85+ displays lower, more significant F_{nd} values than DP SCs (mean F_{nd} = -1.183), but only 65 (48%) populations displayed significant p-values for the 36:85+ pair.

3.4. Randomization Tests of Neutrality

We used DP SC resampling to compare the observed F_{nd} values for DP SCs to the distribution of F_{nd} values that resulted from random assignment of DPB1 alleles to DP

SCs. As presented in Table 5, the true F_{nd} values for DP SCs were significantly lower than the randomized F_{nd} values in 16% of populations. Of these, 91% (20 populations) were European, consistent with the dramatic difference between homozygosity values for *DPB1* alleles and DP SCs illustrated in Figure 2. Overall, the observed F_{nd} values for DP SCs were lower than 50% of randomized values in 94 (69%) populations.

We performed a parallel set of analyses of selection for DP supertypes and T-cell epitopes, but did not infer the action of balancing selection for any of these locus categories.

3.5. Worldwide Distribution of DP serologic categories

The frequency of each DP SC in each population is presented in Supplementary Table S4, and frequency distributions for each DP SC, interpolated for 123 non-migrant populations (identified in Table 5), are illustrated in Figure 1. DP SCs are unevenly distributed in world populations, with each SC common in a different region of the world, and at different frequencies. The DP1 SC is most frequent in Australia, Oceania and Southeast Asia, where 20 of 21 populations with DP1 SC frequencies greater than 0.5 are found. The DP2 SC is most frequent in North and South America; 22 of the 27 populations with DP2 SC frequencies greater than 0.5 are in these regions. The DP3 SC is most frequent in South America. Only four populations have DP3 SC frequencies greater than 0.5, and all are from this region. The DP4 SC is most frequent in European populations. Three of the four populations with DP4 SC frequencies greater than 0.5 are European, as are 24 of the 27 populations with DP4 SC frequencies greater than 0.4. All four DP SCs are present in 117 populations. The DP2 SC is observed in all populations, with a minimum frequency of 0.006 in the Trobriand Islander population (n = 1). Many of the populations missing the DP1, DP3, and DP4 SCs are South American.

An apparently compensatory pattern of frequency distributions between the DP1 and DP4 SCs can be observed in Figure 1. The frequency of the DP1 SC is relatively low in Europe (with a mean frequency of 11%), and high in Southeast Asia (51%), while DP4 SC frequencies are high in Europe (43%) and low in East Asia (13%). The correlation (r) and significance between the frequencies for each pair of DP SCs in non-migrant populations in each region and across all regions is presented in Table 6. When all populations were evaluated without regional distinction, significant negative correlation was observed between DP1 and the DP2, DP3 and DP4 SC frequencies and also between the DP2 and DP4 SCs, after correcting for multiple comparisons. In addition, comparing regional results, significant and very strong negative correlations were observed between the DP1 and DP2 SCs in Europe, Southeast Asia, and Oceania, DP1 and DP4 in Oceania, DP2 and DP3 in South America, and between DP4 and both DP2 and DP3 in Europe.

3.6. DPA1~DP Serologic Category Haplotypes

Estimated *DPA1*~DP SC haplotypes for 10 populations from Sub-Saharan Africa, Europe, Oceania, North America, and South America are presented in Table 7. Although eight DPA1 alleles are observed in the compiled data-set, four alleles (*DPA1*01:03*, **02:01*, **02:02*, and **03:01*) represent the majority of the allelic diversity observed at this locus [24, 111], and these four alleles contribute to the major haplotypes. Although these major

haplotypes are shared across populations, haplotype diversity decreases with distance from Africa. The major DPA1-DP SC haplotypes (frequencies > 0.1 and D'_{ij} > 0.5), are consistent across populations and regions, with the notable exception that *DPA1*03:01* is observed only in Africa. DP1 is associated with *DPA1*02:01* and **02:02*, DP2 with *DPA1*01:03* and *DPA1*03:01*; DP3 with *DPA1*02:01*; and DP4 with *DPA1*01:03*.

The frequencies of *DPA1*01:04*, **02:03*, **03:02*, and **04:01* haplotypes are generally too low to allow interpretation, as rare haplotypes (n = 1 or 2) generated via the EM algorithm are unreliable. However, four *DPA1*04:01*~DP1 haplotypes were observed in both the East Timor and Filipino populations, and the seven DPA1*04:01~DP3 haplotypes observed in the PNG Highland population displayed D'_{ij} > 0.5.

The pattern of LD between these DP loci is determined by the DPB1 residues at positions 85-87. *DPA1*02:01* and **02:02* are associated with the DP1 and DP3 SCs, which share the E85-A86-V87 motif, whereas **01:03* and **03:01* are associated with the DP2 and DP4 SCs, which share the G85-P86-M87 motif. In Oceania, *DPA1*04:01* is associated with DP1 and DP3, and the E85-A86-V87 motif.

For the major haplotypes involving *DPA1*01:03*, **02:01*, **02:02*, and **03:01*, these observations confirm those made by Hollenbach et al. [35] in a large European American population, where DPα AA residue Q31 (present in *DPA1*02:01* and **02:02*) is proposed to interact with the E85-A86-V87 motif, while DPA1 AA residue M31 (present in **01:03* and **03:01*) interacts with the G85-P86-M87 motif, forming stable DP heterodimers. This association is observed even in the South American Ticuna, for which only three *DPA1* alleles are observed. However, *DPA1*04:01* encodes M31 and is associated with the E85-A86-V87 motif. As shown in Table 7, *DPA1*02:01* and **02:02* encode R50 and A83 residues in addition to Q31; R50 and A83 are also encoded by **04:01*.

3.6.1. Collapsing DPA1~DPB1 haplotypes to DPA1~DP SC haplotypes

We compared *DPA1*~DP SC haplotypes generated via the EM algorithm (where DP SCs were determined prior to haplotype estimation) to haplotypes obtained by collapsing the *DPB1* alleles in *DPA1~DPB1* haplotypes to DP SCs (data not shown). Many haplotypes displayed relatively minor differences in haplotype frequency when generated by the two approaches. Of 91 haplotypes estimated in the 10 populations, 55 differed in frequency from those obtained by collapsing; the maximum frequency difference (0.008) resulted from the generation of the *DPA1*0201*~DP2 haplotype in the Slovenian population after collapsing *DPA1~DPB1* haplotypes. These differences between collapsed and estimated haplotypes derive from the unreliability of the EM algorithm in estimating haplotypes that are only observed once or twice; in general, only haplotypes observed at least three times in a population should be given serious consideration. *DPB1* alleles should be collapsed to DP SCs prior to the estimation of haplotypes involving DP SCs.

3.7. Contrasting Patterns of Selection at the Allele- and SC-Levels

The relationship between F_{nd} values calculated in each population for *DPB1* alleles (presented in the companion paper) and the four DP SCs (Table 1) is illustrated in Figures 1 and 2. These values are presented alongside those calculated for T-cell epitopes, DP

STs, and 18 polymorphic *DPB1* exon 2 encoded AA positions in Supplementary Table S5. As previously observed [14, 15, 20][companion paper], *DPB1* allele-frequency distributions are consistent with the null hypothesis of neutral evolution (F_{nd} = 0), with a mean F_{nd} value across all populations of 0.13. Only 3/136 populations display significantly negative F_{nd} for *DPB1* alleles at the 0.05 level (West African, Nu, and Coreguaje).

In contrast, for DP SCs, 73 populations (54%) displayed homozygosity values that were significant at this level, and the mean F_{nd} value for DP SCs is -1.15. As illustrated in Figure 2, this difference between allele-level F_{nd} and DP SC-level F_{nd} values was most pronounced in European populations, where the mean allele-level F_{nd} value is 0.37, while the mean DP SC-level F_{nd} value is -1.62. While no individual p-value is significant after correction for multiple comparisons (all p-values > 0.00037), almost half of the p-values significant at the 0.05 level are for European populations, and 35 of 37 European populations display significant p-values.

While no individual population displays significantly low homozygosity, compared to neutrality expectations, 50% of *DPB1* allele-level F_{nd} values are negative, while 91% of DP SC-level F_{nd} values are negative, a pattern consistent with balancing selection acting on the DP SCs but not on *DPB1* alleles. Parametric two-tailed t-tests of the mean F_{nd} values (versus the null hypothesis of neutral evolution) for *DPB1* alleles and DP SCs yielded similar results, with p-values of 0.11 and 2.9×10^{-39}, respectively. These results provide strong support for the idea that selection is operating differently on *DPB1* alleles and DP SCs.

3.8. Linkage Disequilibrium within DP Serological Categories

We dissected the contributions of DPβ AA variants in each serological category to overall LD by calculating *ALD* across alleles in each set of three serological categories, excluding those in the fourth (e.g. LD for AA pairs in DP2, DP3, and DP4 alleles, with DP1 encoding alleles excluded, is illustrated in Supplementary Figure S1A). This exclusion is possible because *ALD* is calculated from the frequency of individual alleles and their constituent polymorphisms. The results are presented in Supplementary Figure S1A-D. It is clear in this figure that alleles in specific SCs contribute differentially to the LD at the population level. For example, in the absence of DP1 alleles, high mean *ALD* values (ranging from 0.83 to 1.0) are observed between the more N-terminal AA positions (8, 9, and 11) and the more C-terminal positions (84, and 85-87), whereas these values are intermediate (0.64-0.74) when all alleles are considered (Figure 3), and are even lower (0.43-0.71) when other DP serological categories are excluded.

Similarly, *ALD* between positions 36 and 56 is highest in the absence of DP1 encoding alleles ($W_{36|56}$ = 0.9, $W_{56|36}$ = 1.0), suggesting particularly high 36:56 diversity in DP1 encoding alleles. In other cases, *ALD* is differentially impacted by DP serological category status; $W_{56|55}$=1.0 overall and in the absence of alleles from each serological category, while $W_{55|56}$ ranged from a low of 0.76 (DP2 excluded) to 0.97 (DP1 excluded). When DP4 encoding alleles are excluded and position 35 is conditioned on, there is high *ALD* with positions 36, 55, and 56; a pattern not seen when alleles from other serological categories are excluded. In other cases, *ALD* is not impacted by DP serological category:

$W_{33|69}$=1.0 in all cases, while $W_{69|33}$ is low overall (0.31) as well as when each DP category is removed (0.25 – 0.43).

Finally, *ALD* values between positions 36 and 85-87, and 56 and 85-87 are consistently low in all five analyses (ranging from 0.28 to 0.63), a pattern that may be driven by selective pressures inferred to be acting on these pairs.

4. **Discussion**

The nature of selective pressure operating on *DPB1* diversity has been unclear for some time; the ratio of synonymous to non-synonymous substitutions in *DPB1* exon 2 sequences encoding peptide binding residues is consistent with balancing selection [112, 113], revealing ancient instances of strong selection, while DPB1 allele frequencies are consistent with neutral evolution (genetic drift), in the recent past of the human population.

Here and in our companion paper, we have applied frequency-based analyses of selection in a hierarchical fashion, analyzing the frequencies of alleles, individual AA positions, AA pairs and trios, and functionally-defined allele-classes (T-cell epitopes, supertypes, and serological categories), in a large world-wide collection of populations to better characterize the manner in which selection shapes *DPB1* polymorphism. Our approach takes the *frequency* of each variant into account. While 1602 two-field *DPB1* alleles have been defined, only 291 two-field *DPB1* alleles are recognized as being common, intermediate or well-documented [111]. Therefore, a polymorphism that might seem rare in a comparison across allele sequences may be more common in populations as the result of selection.

This approach allows us to characterize the modes of selection operating at the AA level in *DPB1* alleles. Of 87 *DPB1* exon 2 encoded amino acids, 67 are monomorphic, five are evolving under directional selection, four are experiencing genetic drift, and eleven are under balancing selection. In particular, AA positions 36, 55, 56, 84 and 85-87 (this last representing a dimorphic three-residue sequence block) appear to be under strong balancing selection. Among these, the strongest signal of balancing selection is observed for position 56. These results confirm those of Salamon et al. [14] and Valdes et al. [15], which were determined using many fewer populations.

Variants at AA position 56 and the positions 85-87 constitute the primary immunodominant serologic epitopes of the DP molecule. The four DP SCs defined by these epitopes also display evidence of strong balancing selection. This confirms the results of Hollenbach et al.[35] and Voorter et al.[34] which were determined in single populations. This pattern stands in marked contrast to that for *DPB1* alleles (Figure 2). As the DP SCs are functionally defined, we conclude that maintaining population-level diversity in these DP SCs is more evolutionarily relevant than maintaining a repertoire of individual *DPB1* alleles. For example, many populations in Polynesia and the Americas, which are thought to have experienced extreme historical founder effects, have only between five and 10 *DPB1* alleles, but have all four DP SCs (see Supplementary Table S4). Even populations with only four or five *DPB1* alleles have three or four of these SCs represented.

In contrast, analyses of selection on three other functionally defined classes of *DPB1* alleles (TCE3 and TCE4 T-cell epitope groups [27] and DP supertypes [31]) (Supplementary Table S6) reveal a significant trend toward directional selection for TCE3 groups, and a significant trend toward balancing selection for TCE4 groups and DP supertypes. However, the mean homozygosity values for TCE4 T-cell epitopes and DP supertypes (-0.26 and -0.50, respectively) are intermediate relative to other possible locus-categories (Figures 2 and 4 and Supplementary Figure S2), with much lower F_{nd} values for many AA pairs and trios. In terms balancing selection, the DP SCs appear to be fundamentally distinct from other functionally defined allele classes.

Table 8 summarizes the evidence for selection across all populations, and subsets of these populations in seven global regions, for individual alleles, DP SCs, TCE3 TCEs, TCE4 TCEs, and DP STs. Evidence for selection operating on DP SCs is strong across all populations when considered as a group, very strong in populations in Northeast Asia and Europe, strong in Southeast Asian populations and moderate in Sub-Saharan African populations. Evidence for selection operating on DP STs is strong in Southeast Asian populations as well. Evidence of selection on individual DPB1 alleles and TCEs is consistently weak across all populations in all regions.

4.1. Balancing Selection at DPB1 is Not Limited to Peptide-Binding Positions

Residues at positions 56 and 86-87, which define the DP SCs, point away from the peptide binding groove, and position 56 is outside of the TCR footprint (see Figure 5 of the companion paper). However, positions 85-87 contribute to the contact area for the DP α and β chains. It remains unclear why the primary immunodominant DP serologic epitopes should be under such strong balancing selection, when peptide and TCR interactions are minimal. For example, given the very high LD between 55 and 56 and 84 and 85-87 (Figure 3), it might be the case that selection is primarily operating on positions 55 and 84, and that the variants at position 56 and in the 85-87 block are hitchhiking with their peptide-binding neighbors.

Three observations argue against the notion of pure hitchhiking for positions 36, 56, and 85-87. First, as illustrated in Figure 6, F_{nd} values for AA pairs involving these positions, are distinctly lower than for AA pairs that do not involve any of these three positions. This distinction is highly significant when comparing AA pairs involving one or more of positions 36, 56, or 85-87 to pairs that do not involve any of these positions (p-value < 0.00001). This suggests that the low homozygosity at positions 56 and 85-87 is due to more than proximity to positions 55 and 84.

A second argument against purely hitchhiking for positions 36, 56, and 85-87 is that previously identified connections between DPα position 31 and DPβ positions 85-87 in determining *DPA1~DPB1* haplotype associations (Hollenbach et al. [35]) appear to be more complex when considered in multiple non-European populations. We observed the E85-A86-V87 residue block in LD with *DPA1*02:01* (encoding Q31), **02:02* (Q31) *and *04:01* (M31), and the G85-P86-M87 residue block in LD with **01:03* (M31) and **03:01* (M31) in multiple populations around the world (Table 7). Variation at position 50 may

alter the contact region between the DPα and DPβ chains, and we suggest that DPα position 50 is also important for DP dimer formation. Table 7 shows that position R50 vs. Q50 aligns with these LD blocks. Like DPα position 31, position 50 is also in close proximity to DPβ E85-A86-V87, while position 83 is in the α2 domain (see Figure 5 of the companion paper). While position 50 is outside of the peptide binding groove, Lauterbach et al. recently applied structural modeling to conclude that DPα positions 31 and 50 variation may influence peptide binding and TCR recognition[114]. Additional detail about this is included in the supplementary material.

The third argument against purely hitchhiking for positions 36, 56, and 85-87 is that position 36 displays no hitchhiking influence on its neighbors. For example, as shown in Table 3, position 35 is polymorphic but displays frequencies that are consistent with neutral evolution. As illustrated in Figure 3, LD between positions 35 and 36 is intermediate, although this is primarily due to low LD between these positions in alleles in the DP4 SC (Supplementary Figure S1D). The strong balancing selection observed for position 36 also extends to AA pairs, and balancing selection on the 36:85+ pair is equivalent to that for DP SCs. However, polymorphism at position 36 appears to be constrained depending on the sequence at position 56, with V36 residues favored in molecules that have E56 residues (Supplementary Table S7). This constraint suggests a specific functional role for V36:E56 molecules, and may explain why alleles in the DP2 SC are observed in every population. This constraint also explains why the strong balancing selection observed for positions 36, 56, and 85-87 does not extend to the 36:56:85+ trio; trios involving A36:E56 are relatively rare, with a mean frequency of only 0.05.

Despite these observations, and the finding that position 56 variation influences the structure of the DPβ α-helical domain [25], the reason for the strong balancing selection at position 56 remains unclear. The position 85-87 block plays a key role in DP function, and it seems likely that position 56 does as well.

4.2 The DP Serologic Categories Appear to be Under the Strongest Balancing Selection in Europe

As shown in Table 8, we observed consistent differences between the strength of balancing selection in European and non-European populations. The difference in normalized homozygosity measures between *DPB1* alleles and DP SCs is most pronounced in European populations, as is the difference between observed DP SCs and those assigned by the permuted assignment of alleles to equivalently sized categories (Figure 1). The significant negative correlation between the frequencies of the DP1 and DP2 SCs in Europe, Southeast Asia, and Oceania is also most pronounced in Europe. It is not clear why European populations should be distinguished in this manner. One explanation might be a potential ascertainment bias in the genotyping of *DPB1* alleles, given that many common alleles were first identified in European populations. Such a bias could result in a false conclusion of balancing selection when it results in the failure to detect low-frequency alleles. However, DP SCs are not the specific targets of any DPB1 genotyping system; given that they reflect the sequence of two routinely genotyped DPβ AA positions, the basis for such a potential bias is not clear. Alternatively, natural selection

favoring DP SC diversity may genuinely be highest in Europe. Resolution of this issue may have to wait until a better understanding of the functional role of this intriguing category of DPB1 polymorphism is available.

5. Conclusions

We have performed a comprehensive analysis of selection on the functional categories of DPB1 polymorphism, defined both by functional subsets of DPB1 alleles (T-Cell Epitopes) and by key DPB1 AA positions (Supertypes and Serologic Categories), as well as other possible combinations of DPB1 AA positions. We conclude that, as has been identified for other HLA loci, the DPB1 locus is also evolving under Balancing Selection, but that this selection operates primarily at the level of DP Serologic Categories, suggesting a mode of evolution at the DPB1 locus that is distinct from that of other HLA genes.

We have shown that the strength of the balancing selection for DPβ AA positions 36, 56, and 85-87 does not extend beyond the level of AA pairs, and that patterns of LD are distinct within each DP SC, reflecting functional constraints on the overall sequence of DPβ molecules. Natural selection at the *HLA-DPB1* locus favors the presence of key variants at a very small number of AA positions. So long as these key variants are present in a population, a variety of combinations of variants at other AA positions are permitted; genetic drift allows population differentiation of frequent *DPB1* alleles, even while balancing selection acts to maintain the diversity of DP SCs. As a result, the *DPB1* locus displays the highest degree of population-level differentiation among HLA loci [21].

Selection for diversity at individual AA positions appears to be counterbalanced by functional constraints on the sequence of the DPβ molecule. The selection enforcing these specific constraints may be relatively recent in the history of the human species. This explains why balancing selection cannot be inferred from population analyzes of *DPB1* allele frequencies, whereas it can be inferred from analyses of *DPB1* nucleotide sequences. The extension of these analyses to other *HLA* loci will determine the extent to which this phenomenon is specific to the *DPB1* locus.

Acknowledgements

This work was supported by National Institutes of Health (NIH) grants R01AI029042 (SJM, HAE) and R01AI128775 (SJM) awarded by the National Institute of Allergy and Infectious Diseases (NIAID), NIH Contract HHSN272201200028C (RMS and GT) and a REACH Grant from the University of Vermont (RMS). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Allergy and Infectious Diseases, NIH or the United States Government. We thank Dr. Jill A. Hollenbach for data access and helpful discussions. No artificial intelligence systems were used in the writing of this paper or for the work described.

Literature Cited

1. Begovich, A.B., et al., *Polymorphism, recombination and linkage disequilibrium within the HLA class II region.* J. Immunol., 1992. **148**: p. 249.
2. Bugawan, T.L., et al., *High-resolution HLA class I typing in the CEPH families: Analysis of linkage disequilibrium among HLA loci.* Tissue Antigens, 2000. **56**(5): p. 392-404.
3. Sasazuki, T., et al., *Gene Map of the HLA Region, Graves' Disease and Hashimoto Thyroiditis, and Hematopoietic Stem Cell Transplantation.* Adv Immunol, 2016. **129**: p. 175-249.
4. Hildesheim, A., et al., *Association of HLA class I and II alleles and extended haplotypes with nasopharyngeal carcinoma in Taiwan.* J Natl Cancer Inst, 2002. **94**(23): p. 1780-9.
5. Stewart, C.A., et al., *Complete MHC haplotype sequencing for common disease gene mapping.* Genome Res, 2004. **14**(6): p. 1176-87.
6. Aly, T.A., et al., *Extreme genetic risk for type 1A diabetes.* Proc Natl Acad Sci U S A, 2006. **103**(38): p. 14074-9.
7. de Bakker, P.I., et al., *A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC.* Nat Genet, 2006. **38**(10): p. 1166-1172.
8. Thomson, G., et al., *Relative predispositional effects of HLA class II DRB1-DQB1 haplotypes and genotypes on type 1 diabetes: a meta-analysis.* Tissue Antigens, 2007. **70**(2): p. 110-27.
9. Yamazaki, A., et al., *Human leukocyte antigen class I polymorphisms influence the mild clinical manifestation of Plasmodium falciparum infection in Ghanaian children.* Hum Immunol, 2011. **72**(10): p. 881-8.
10. Morris, D.L., et al., *Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans.* Am J Hum Genet, 2012. **91**(5): p. 778-93.
11. Apps, R., et al., *Influence of HLA-C expression level on HIV control.* Science, 2013. **340**(6128): p. 87-91.
12. Meyer, D. and G. Thomson, *How selection shapes variation of the human major histocompatibility complex: A review.* Ann Hum Genet, 2001. **65**(Pt 1): p. 1-26.
13. Hedrick, P.W. and G. Thomson, *Evidence for balancing selection at HLA.* Genetics, 1983. **104**(3): p. 449-56.
14. Salamon, H., et al., *Evolution of HLA class II molecules: Allelic and amino acid site variability across populations.* Genetics, 1999. **152**: p. 393-400.
15. Valdes, A.M., et al., *Locus and population specific evolution in HLA class II genes.* Annals of Human Genetics, 1999. **63**: p. 27-43.
16. Mack, S.J., et al., *Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions.* Tissue Antigens, 2000. **55**(5): p. 383-400.
17. Meyer, D., et al., *Signatures of demographic history and natural selection in the human major histocompatibility complex Loci.* Genetics, 2006. **173**(4): p. 2121-42.
18. Meyer, D., et al., *Single locus polymorphism of classical HLA genes*, in *Immunobiology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Conference. Vol 1*, J.A. Hansen, Editor. 2007, IHWG Press: Seattle, WA. p. 653-704.
19. Tsai, Y. and G. Thomson, *Selection intensity differences in seven HLA loci in many populations*, in *Immunobiology of the Human MHC. Proceedings of the 13th International Histocompatibility Workshop and Conference*, J.A. Hansen, Editor. 2007, IHWG Press: Seattle, WA. p. 199-201.
20. Solberg, O.D., et al., *Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies.* Hum Immunol, 2008. **69**(7): p. 443-64.

21. Buhler, S. and A. Sanchez-Mazas, *HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events.* PLoS One, 2011. **6**(2): p. e14643.
22. Riccio, M.E., et al., *16(th) IHIW: analysis of HLA population data, with updated results for 1996 to 2012 workshop data (AHPD project report).* Int J Immunogenet, 2013. **40**(1): p. 21-30.
23. Begovich, A.B., et al., *Genetic variability and linkage disequilibrium within the HLA-DP region: analysis of 15 different populations.* Tissue Antigens, 2001. **57**(5): p. 424-39.
24. Sanchez-Mazas, A., et al., *Immunogenetics as a tool in anthropological studies.* Immunology, 2011. **133**(2): p. 143-64.
25. Diaz, G., et al., *Functional analysis of HLA-DP polymorphism: a crucial role for DPbeta residues 9, 11, 35, 55, 56, 69 and 84-87 in T cell allorecognition and peptide binding.* Int Immunol, 2003. **15**(5): p. 565-76.
26. Diaz, G., et al., *HLA-DPbeta residue 69 plays a crucial role in allorecognition.* Tissue Antigens, 1998. **52**(1): p. 27-36.
27. Sizzano, F., et al., *Significantly higher frequencies of alloreactive CD4+ T cells responding to nonpermissive than to permissive HLA-DPB1 T-cell epitope disparities.* Blood, 2010. **116**(11): p. 1991-2.
28. Zino, E., et al., *Frequency and targeted detection of HLA-DPB1 T cell epitope disparities relevant in unrelated hematopoietic stem cell transplantation.* Biol Blood Marrow Transplant, 2007. **13**(9): p. 1031-40.
29. Zino, E., et al., *A T-cell epitope encoded by a subset of HLA-DPB1 alleles determines nonpermissive mismatches for hematologic stem cell transplantation.* Blood, 2004. **103**(4): p. 1417-24.
30. Crocchiolo, R., et al., *Nonpermissive HLA-DPB1 disparity is a significant independent risk factor for mortality after unrelated hematopoietic stem cell transplantation.* Blood, 2009. **114**(7): p. 1437-44.
31. Taylor, G.M., et al., *HLA-associated susceptibility to childhood B-cell precursor ALL: definition and role of HLA-DPB1 supertypes.* Br J Cancer, 2008. **98**(6): p. 1125-31.
32. Taylor, G.M., et al., *Relationship between HLA-DP supertype and survival in childhood acute lymphoblastic leukaemia: evidence for selective loss of immunological control of residual disease?* Br J Haematol, 2009. **145**(1): p. 87-95.
33. Cano, P. and M. Fernandez-Vina, *Two sequence dimorphisms of DPB1 define the immunodominant serologic epitopes of HLA-DP.* Hum Immunol, 2009. **70**(10): p. 836-43.
34. Voorter, C.E.M., et al., *Allele and haplotype frequencies of HLA-DPA1 and -DPB1 in the population of Guadeloupe.* Tissue Antigens, 2014. **83**(3): p. 147-153.
35. Hollenbach, J.A., et al., *A combined DPA1~DPB1 amino acid epitope is the primary unit of selection on the HLA-DP heterodimer.* Immunogenetics, 2012. **64**(8): p. 559-69.
36. Renquin, J., et al., *HLA class II polymorphism in Aka Pygmies and Bantu Congolese and a reassessment of HLA-DRB1 African diversity.* Tissue Antigens, 2001. **58**(4): p. 211-22.
37. Gonzalez-Galarza, F.F., et al., *Allele frequency net: a database and online repository for immune gene frequencies in worldwide populations.* Nucleic Acids Res, 2011. **39**(Database issue): p. D913-9.
38. May, J., et al., *HLA DPA1/DPB1 genotype and haplotype frequencies, and linkage disequilibria in Nigeria, Liberia, and Gabon.* Tissue Antigens, 1998. **52**(3): p. 199-207.
39. Mack, S.J., et al., *Anthropology/human genetic diversity population reports*, in *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference*, J. Hansen, Editor. 2007, IHWG Press: Seattle. p. 580-652.
40. Magzoub, M.M., et al., *HLA-DP polymorphism in Sudanese controls and patients with insulin-dependent diabetes mellitus.* Tissue Antigens, 1992. **40**(2): p. 64-8.

41. Aldener-Cannava, A. and O. Olerup, *HLA-DPB1 typing by polymerase chain reaction amplification with sequence-specific primers.* Tissue Antigens, 2001. **57**(4): p. 287-99.
42. Hmida, S., et al., *HLA class II gene polymorphism in Tunisians.* Tissue Antigens, 1995. **45**(1): p. 63-8.
43. Ayed, K., et al., *HLA class-I and HLA class-II phenotypic, gene and haplotypic frequencies in Tunisians by using molecular typing data.* Tissue Antigens, 2004. **64**(4): p. 520-32.
44. Lienert, K., et al., *HLA DPB1 genotyping in Australian aborigines by amplified fragment length polymorphism analysis.* Hum Immunol, 1993. **36**(3): p. 137-41.
45. Pickl, W.F., I. Fae, and G.F. Fischer, *Detection of established and novel alleles of the HLA-DPB1 locus by PCR-SSO.* Vox Sang, 1993. **65**(4): p. 316-9.
46. Comas, D., et al., *HLA class I and class II DNA typing and the origin of Basques.* Tissue Antigens, 1998. **51**(1): p. 30-40.
47. Perez-Miranda, A.M., et al., *Genetic polymorphism and linkage disequilibrium of the HLA-DP region in Basques from Navarre (Spain).* Tissue Antigens, 2004. **64**(3): p. 264-75.
48. Raguenes, O., et al., *HLA class II typing and idiopathic IgA nephropathy (IgAN): DQB1*0301, a possible marker of unfavorable outcome.* Tissue Antigens, 1995. **45**(4): p. 246-9.
49. Sage, D.A., P.R. Evans, and W.M. Howell, *HLA DPA1-DPB1 linkage disequilibrium in the British caucasoid population.* Tissue Antigens, 1994. **44**(5): p. 335-8.
50. Wu, Z., et al., *Molecular analysis of HLA-DQ and -DP genes in caucasoid patients with Hashimoto's thyroiditis.* Tissue Antigens, 1994. **43**(2): p. 116-9.
51. Perdriger, A., et al., *DPB1 polymorphism in rheumatoid arthritis: evidence of an association with allele DPB1 0401.* Tissue Antigens, 1992. **39**(1): p. 14-8.
52. Begovich, A.B., et al., *Genes within the HLA class II region confer both predisposition and resistance to primary biliary cirrhosis.* Tissue Antigens, 1994. **43**(2): p. 71-7.
53. Vambergue, A., et al., *Gestational diabetes mellitus and HLA class II (-DQ, -DR) association: The Digest Study.* Eur J Immunogenet, 1997. **24**(5): p. 385-94.
54. Begovich, A.B., et al., *Polymorphism, recombination, and linkage disequilibrium within the HLA class II region.* J Immunol, 1992. **148**(1): p. 249-58.
55. Hviid, T.V., H.O. Madsen, and N. Morling, *HLA-DPB1 typing with polymerase chain reaction and restriction fragment length polymorphism technique in Danes.* Tissue Antigens, 1992. **40**(3): p. 140-4.
56. Sage, D.A., et al., *HLA DPB1 alleles and susceptibility to rheumatoid arthritis.* Eur J Immunogenet, 1991. **18**(4): p. 259-63.
57. al-Daccak, R., et al., *Gene polymorphism of HLA-DPB1 and DPA1 loci in caucasoid population: frequencies and DPB1-DPA1 associations.* Hum Immunol, 1991. **31**(4): p. 277-85.
58. Bera, O., et al., *HLA class I and class II allele and haplotype diversity in Martinicans.* Tissue Antigens, 2001. **57**(3): p. 200-7.
59. Yao, Z., et al., *DNA typing for HLA-DPB1-alleles in German patients with systemic lupus erythematosus using the polymerase chain reaction and DIG-ddUTP-labelled oligonucleotide probes. Members of SLE Study Group.* Eur J Immunogenet, 1993. **20**(4): p. 259-66.
60. Pratsidou-Gertsi, P., et al., *Nationwide collaborative study of HLA class II associations with distinct types of juvenile chronic arthritis (JCA) in Greece.* Eur J Immunogenet, 1999. **26**(4): p. 299-310.
61. Papassavas, E.C., et al., *MHC class I and class II phenotype, gene, and haplotype frequencies in Greeks using molecular typing data.* Hum Immunol, 2000. **61**(6): p. 615-23.
62. Reveille, J.D., et al., *HLA-class II alleles and C4 null genes in Greeks with systemic lupus erythematosus.* Tissue Antigens, 1995. **46**(5): p. 417-21.

63. Mazzola, G., et al., *Immunoglobulin and HLA-DP genes contribute to the susceptibility to juvenile dermatitis herpetiformis.* Eur J Immunogenet, 1992. **19**(3): p. 129-39.
64. Savage, D.A., et al., *Frequency of HLA-DPB1 alleles, including a novel DPB1 sequence, in the Northern Ireland population.* Hum Immunol, 1992. **33**(4): p. 235-42.
65. Spurkland, A., et al., *Susceptibility to develop celiac disease is primarily associated with HLA-DQ alleles.* Hum Immunol, 1990. **29**(3): p. 157-65.
66. Congia, M., et al., *A high frequency of the A30, B18, DR3, DRw52, DQw2 extended haplotype in Sardinian celiac disease patients: further evidence that disease susceptibility is conferred by DQ A1*0501, B1*0201.* Tissue Antigens, 1992. **39**(2): p. 78-83.
67. Kapustin, S., et al., *HLA class II molecular polymorphisms in healthy Slavic individuals from North-Western Russia.* Tissue Antigens, 1999. **54**(5): p. 517-20.
68. Cechova, E., et al., *HLA-DRB1, -DQB1 and -DPB1 polymorphism in the Slovak population.* Tissue Antigens, 1998. **51**(5): p. 574-6.
69. Sanchez-Velasco, P. and F. Leyva-Cobian, *The HLA class I and class II allele frequencies studied at the DNA level in the Svanetian population (Upper Caucasus) and their relationships to Western European populations.* Tissue Antigens, 2001. **58**(4): p. 223-33.
70. Allen, M., et al., *Association of susceptibility to multiple sclerosis in Sweden with HLA class II DRB1 and DQB1 alleles.* Hum Immunol, 1994. **39**(1): p. 41-8.
71. Sawitzke, A.D., A.L. Sawitzke, and R.H. Ward, *HLA-DPB typing using co-digestion of amplified fragments allows efficient identification of heterozygous genotypes.* Tissue Antigens, 1992. **40**(4): p. 175-81.
72. Rossman, M.D., et al., *HLA-DRB1*1101: a significant risk factor for sarcoidosis in blacks and whites.* Am J Hum Genet, 2003. **73**(4): p. 720-35.
73. Al-Hussein, K.A., et al., *HLA class II sequence-based typing in normal Saudi individuals.* Tissue Antigens, 2002. **60**(3): p. 259-61.
74. Gao, X.J., et al., *DNA typing for HLA-DR, and -DP alleles in a Chinese population using the polymerase chain reaction (PCR) and oligonucleotide probes.* Tissue Antigens, 1991. **38**(1): p. 24-30.
75. Hu, W.H., et al., *Polymorphism of the DPB1 locus in Hani ethnic group of south-western China.* Int J Immunogenet, 2005. **32**(6): p. 421-3.
76. Lin, J.H., et al., *Molecular analyses of HLA-DRB1, -DPB1, and -DQB1 in Jing ethnic minority of Southwest China.* Hum Immunol, 2003. **64**(8): p. 830-4.
77. Chen, S., et al., *Origin of Tibeto-Burman speakers: evidence from HLA allele distribution in Lisu and Nu inhabiting Yunnan of China.* Hum Immunol, 2007. **68**(6): p. 550-9.
78. Geng, L., et al., *Determination of HLA class II alleles by genotyping in a Manchu population in the northern part of China and its relationship with Han and Japanese populations.* Tissue Antigens, 1995. **46**(2): p. 111-6.
79. Liu, Y., et al., *Polymorphism of HLA class II genes in Miao and Yao nationalities of Southwest China.* Tissue Antigens, 2006. **67**(2): p. 157-9.
80. Fu, Y., et al., *HLA-DRB1, DQB1 and DPB1 polymorphism in the Naxi ethnic group of South-western China.* Tissue Antigens, 2003. **61**(2): p. 179-83.
81. Hu, W., et al., *Sequencing-based analysis of the HLA-DPB1 polymorphism in Nu ethnic group of south-west China.* Int J Immunogenet, 2006. **33**(6): p. 397-400.
82. Liu, Z.H., et al., *HLA-DPB1 allelic frequency of the Pumi ethnic group in south-west China and evolutionary relationship of Pumi with other populations.* Eur J Immunogenet, 2002. **29**(3): p. 259-61.
83. Zhou, L., et al., *Polymorphism of human leukocyte antigen-DRB1, -DQB1, and -DPB1 genes of Shandong Han population in China.* Tissue Antigens, 2005. **66**(1): p. 37-43.

84. Wang, F.Q., et al., *HLA-DP distribution in Shanghai Chinese--a study by polymerase chain reaction--restriction fragment length polymorphism.* Hum Immunol, 1992. **33**(2): p. 129-32.
85. Zimdahl, H., et al., *Towards understanding the origin and dispersal of Austronesians in the Solomon Sea: HLA class II polymorphism in eight distinct populations of Asia-Oceania.* Eur J Immunogenet, 1999. **26**(6): p. 405-16.
86. Velickovic, Z.M. and J.M. Carter, *HLA-DPA1 and DPB1 polymorphism in four Pacific Islands populations determined by sequencing based typing.* Tissue Antigens, 2001. **57**(6): p. 493-501.
87. Bugawan, T.L., et al., *PCR/oligonucleotide probe typing of HLA class II alleles in a Filipino population reveals an unusual distribution of HLA haplotypes.* Am J Hum Genet, 1994. **54**(2): p. 331-40.
88. Tracey, M.C. and J.M. Carter, *Class II HLA allele polymorphism: DRB1, DQB1 and DPB1 alleles and haplotypes in the New Zealand Maori population.* Tissue Antigens, 2006. **68**(4): p. 297-302.
89. Mitsunaga, S., et al., *Family study on HLA-DPB1 polymorphism: linkage analysis with HLA-DR/DQ and two "new" alleles.* Hum Immunol, 1992. **34**(3): p. 203-11.
90. Ohta, H., et al., *Histocompatibility antigens and alleles in Japanese haemophilia A patients with or without factor VIII antibodies.* Tissue Antigens, 1999. **54**(1): p. 91-7.
91. Munkhbat, B., et al., *Molecular analysis of HLA polymorphism in Khoton-Mongolians.* Tissue Antigens, 1997. **50**(2): p. 124-34.
92. Hollenbach, J.A., et al., *HLA diversity, differentiation, and haplotype evolution in Mesoamerican Natives.* Hum Immunol, 2001. **62**(4): p. 378-90.
93. Briceno, I., et al., *HLA-DPB1 polymorphism in seven South American Indian tribes in Colombia.* Eur J Immunogenet, 1996. **23**(3): p. 235-40.
94. Gendzekhadze, K., et al., *HLA-DP polymorphism in Venezuelan Amerindians.* Hum Immunol, 2004. **65**(12): p. 1483-8.
95. Cerna, M., et al., *Differences in HLA class II alleles of isolated South American Indian populations from Brazil and Argentina.* Hum Immunol, 1993. **37**(4): p. 213-20.
96. Vullo, C.M., et al., *HLA polymorphism in a Mataco South American Indian tribe: serology of class I and II antigens. Molecular analysis of class II polymorphic variants.* Hum Immunol, 1992. **35**(4): p. 209-14.
97. Layrisse, Z., et al., *Extended HLA haplotypes in a Carib Amerindian population: the Yucpa of the Perija Range.* Hum Immunol, 2001. **62**(9): p. 992-1000.
98. Just, J.J., et al., *African-American HLA class II allele and haplotype diversity.* Tissue Antigens, 1997. **49**(5): p. 547-55.
99. Erlich, H.A., et al., *Association of HLA-DPB1*0301 with IDDM in Mexican-Americans.* Diabetes, 1996. **45**(5): p. 610-4.
100. Lancaster, A., et al., *PyPop: a software framework for population genomics: analyzing large-scale multi-locus genotype data.* Pac Symp Biocomput, 2003: p. 514-25.
101. Lancaster, A.K., et al., *PyPop update - a software pipeline for large-scale multi-locus population genomics.* Tissue Antigens, 2007. **69**: p. 192-197.
102. Thomson, G. and R.M. Single, *Conditional asymmetric linkage disequilibrium (ALD): extending the biallelic r2 measure.* Genetics, 2014. **198**(1): p. 321-31.
103. Single, R.M., et al., *Asymmetric linkage disequilibrium: Tools for assessing multiallelic LD.* Hum Immunol, 2016. **In Press**.
104. Wessel, P. and W.H.F. Smith, *New, improved version of generic mapping tools released.* Eos Trans AGU, 1988. **79**: p. 579.
105. Dai, S., et al., *Crystal structure of HLA-DP2 and implications for chronic beryllium disease.* Proc Natl Acad Sci U S A, 2010. **107**(16): p. 7425-30.

106. Wang, Y., et al., *Cn3D: sequence and structure views for Entrez.* Trends Biochem Sci, 2000. **25**(6): p. 300-2.
107. Mack, S.J. and J.A. Hollenbach, *Allele Name Translation Tool and Update NomenCLature: software tools for the automated translation of HLA allele names between successive nomenclatures.* Tissue Antigens, 2010. **75**(5): p. 457-61.
108. Mack, S.J., et al., *Human leukocyte antigen-A, -B, -C, -DRB1 allele and haplotype frequencies in Americans originating from southern Europe: contrasting patterns of population differentiation between Italian and Spanish Americans.* Hum Immunol, 2011. **72**(2): p. 144-9.
109. Cramer, H., *Mathematical methods of statistics*. 1946, Princeton, NJ: Princeton University Press.
110. Lancaster, A., *Identifying associations between natural selection and molecular function in human MHC genes. Ph.D. Thesis*, in *Integrative Biology*. 2006, University of California, Berkeley: Berkeley, CA. p. 149.
111. Mack, S.J., et al., *Common and well-documented HLA alleles: 2012 update to the CWD catalogue.* Tissue Antigens, 2013. **81**(4): p. 194-203.
112. Hughes, A.L. and M. Nei, *Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection.* Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(3): p. 958-62.
113. Hughes, A.L. and M. Yeager, *Natural selection and the evolutionary history of major histocompatibility complex loci.* Frontiers in Bioscience, 1998. **3**: p. D509-516.
114. Lauterbach, N., et al., *Allorecognition of HLA-DP by CD4+ T cells is affected by polymorphism in its alpha chain.* Mol Immunol, 2014. **59**(1): p. 19-29.

Figure 1. Heat Maps Depicting Interpolated DP1-DP4 Serologic Category Frequency Distributions.

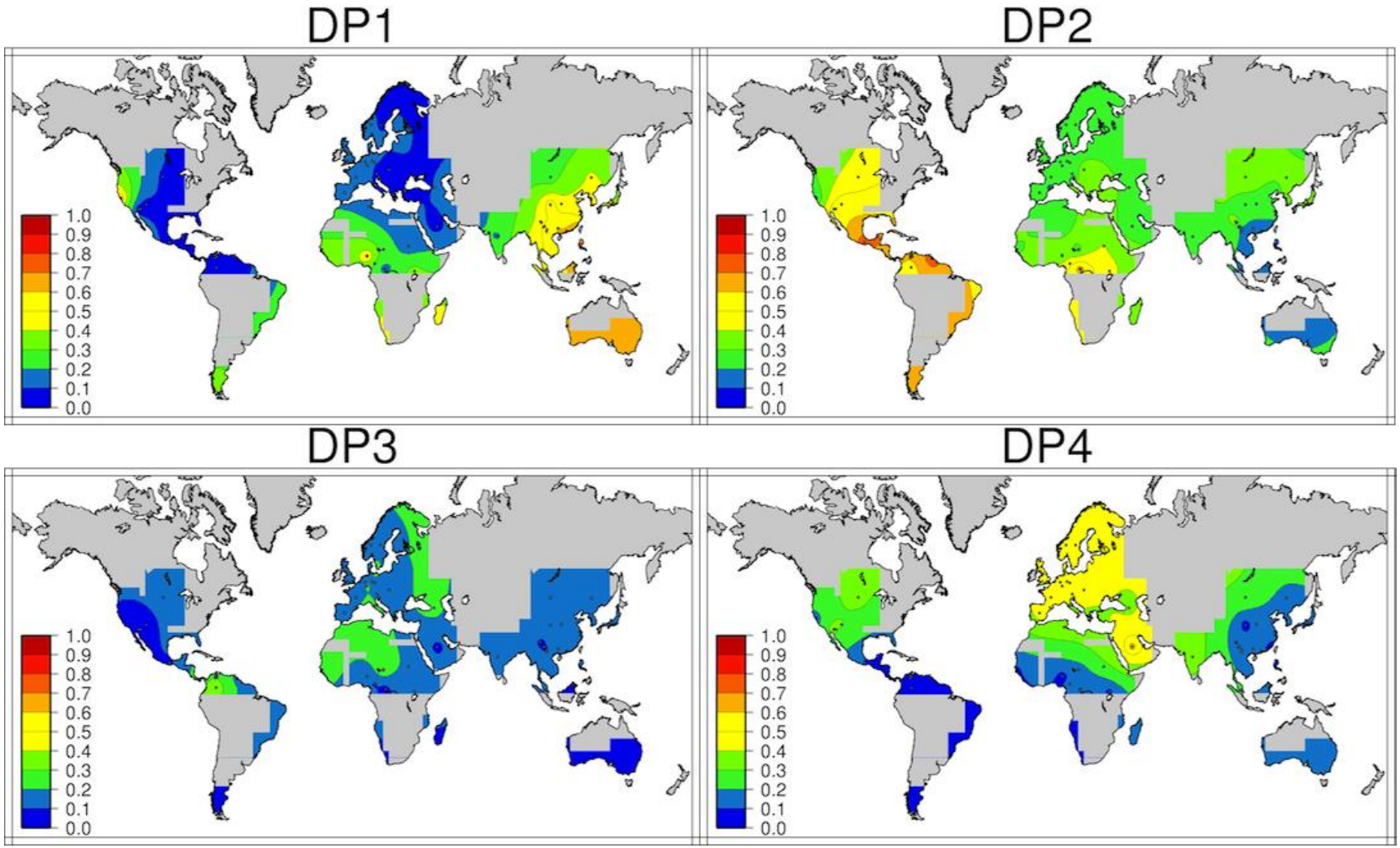

Figure 2. F_{nd} Values for *DPB1* Alleles (□) and DP Serologic Categories (◊)

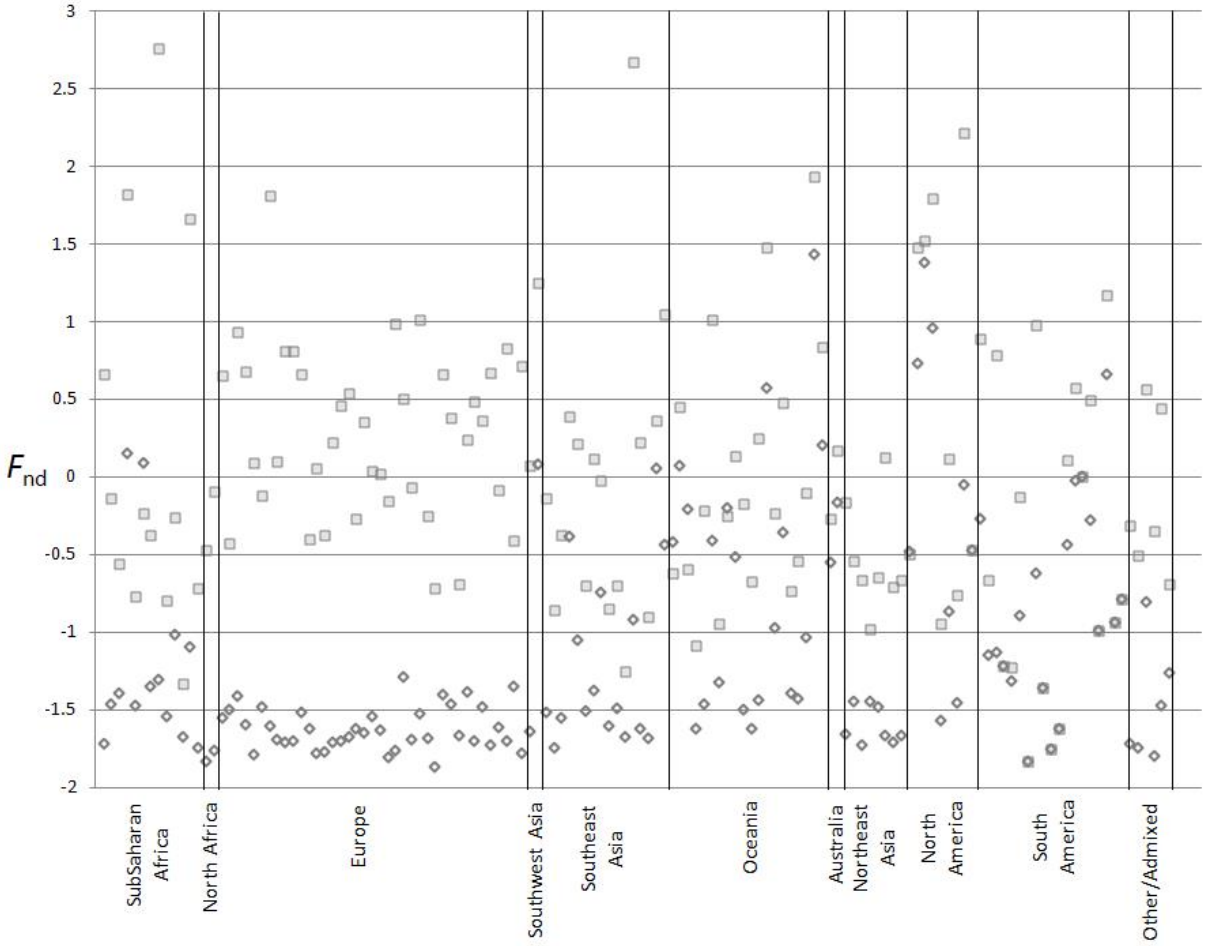

Figure 3. Mean ALD Values for 91 Pairs of DPB1 Encoded Amino Acid Positions (row conditioned on column)

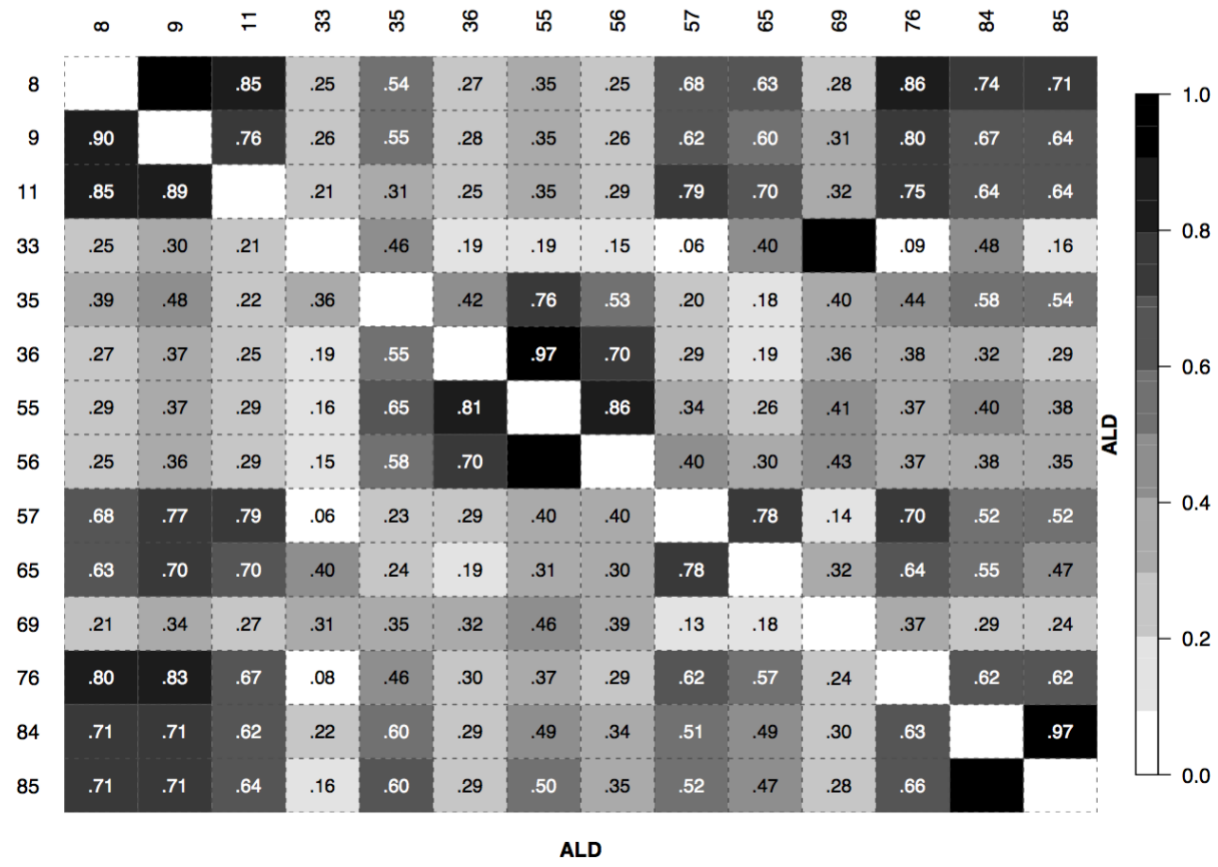

Figure 4. Mean F_{nd} Values For Pairs of Variant *DPB1* Exon 2 Amino Acid Positions

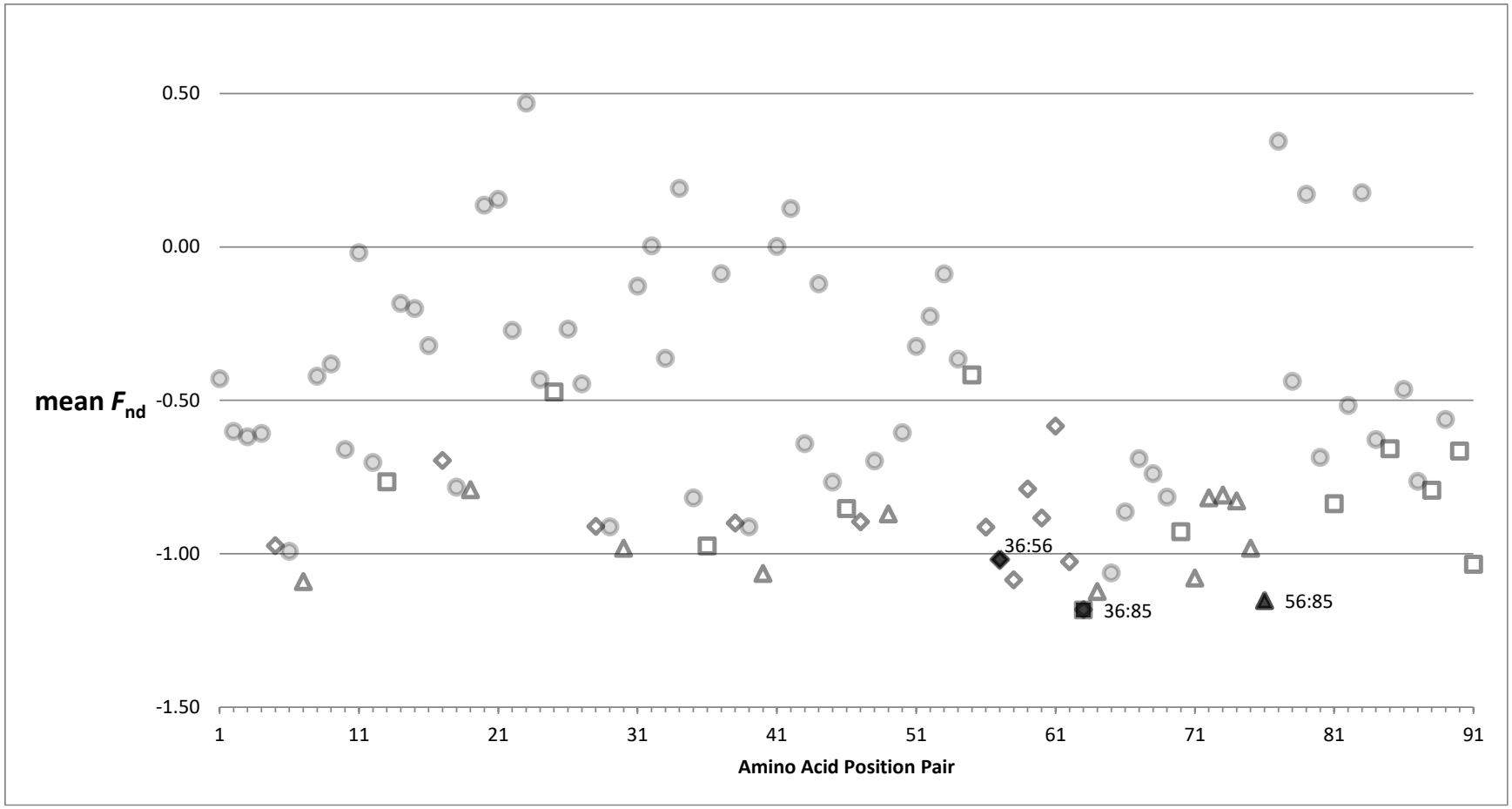

Figure 5. Mean F_{nd} Values for Six AA Pairs and their Constituent AA Positions

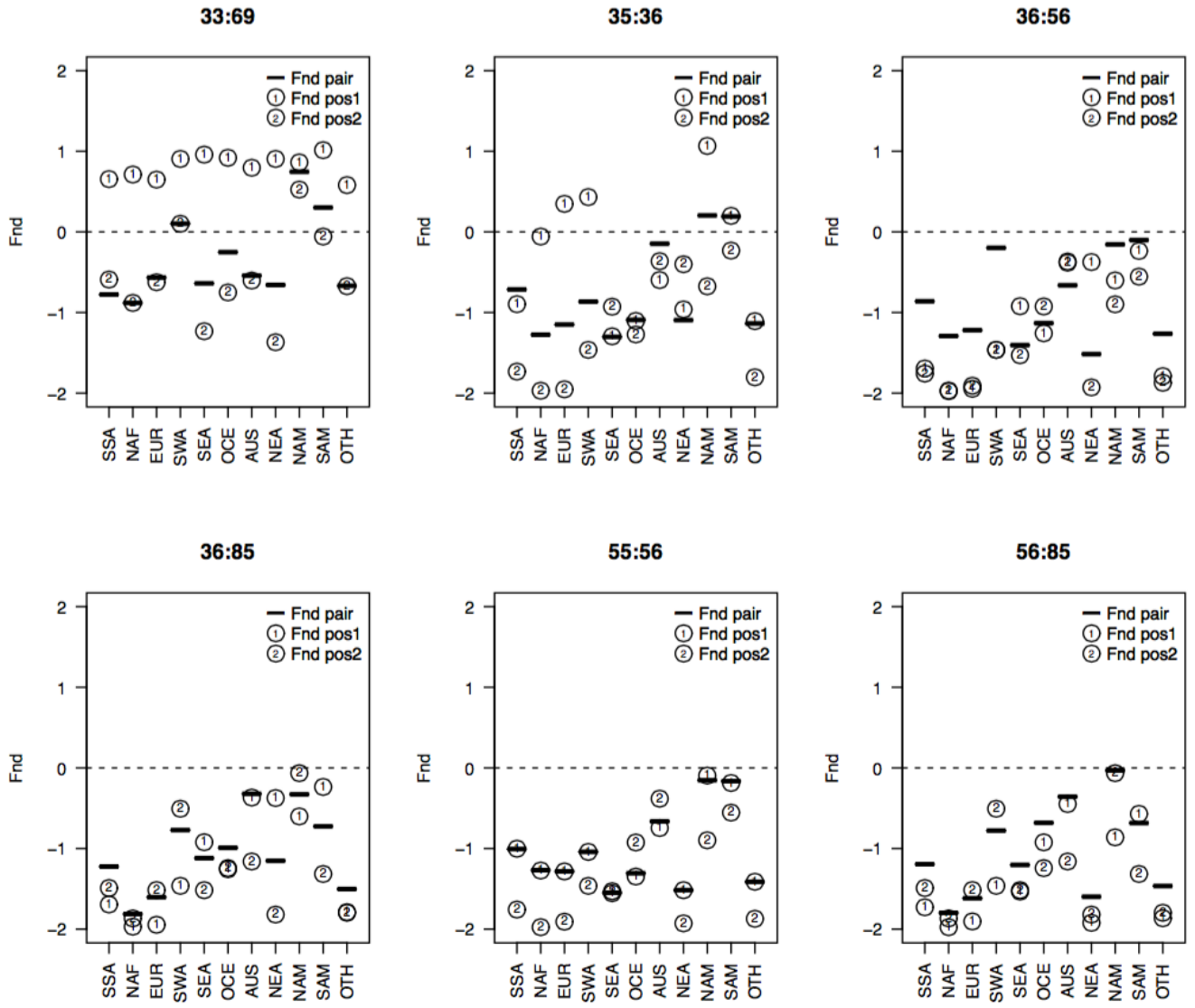

Figure 6. Distribution of mean F_{nd} values by count of polymorphic DPB1 AA pairs containing positions 36, 56, and 85+

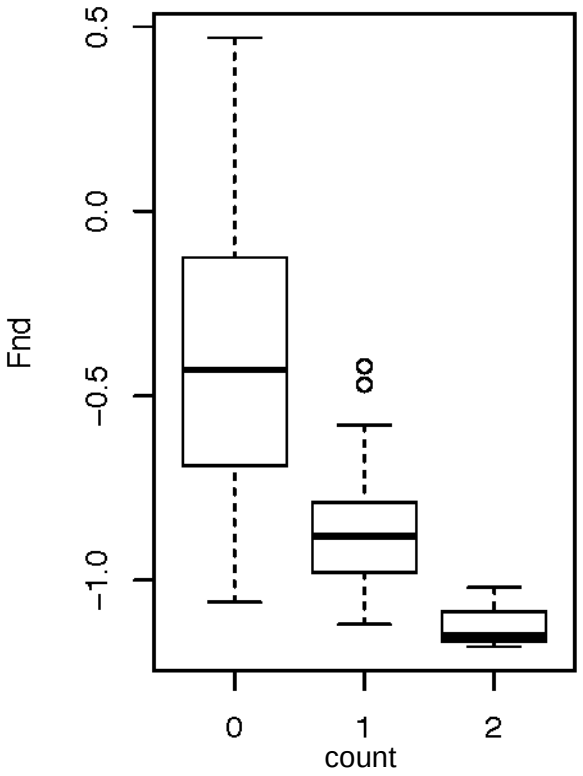

Table 1. Functionally Defined Categories of DPB1 Alleles

DPB1 Functional Category	Subcategory	Definition
TCE3 T-cell epitopes		
	TCE3 group 1	DPB1*09:01, *10:01, *17:01
	TCE3 group 2	DPB1*03:01, *14:01, *45:01 (*86:01[a], *104:01[a])
	TCE3 group 3	All other DPB1 alleles
TCE4 T-cell epitopes		
	TCE4 group 1	DPB1*09:01, *10:01, *17:01
	TCE4 group 2	DPB1*03:01, *14:01, *45:01
	TCE4 group 3	DPB1*02:01, *02:02
	TCE4 group 4	All other DPB1 alleles
DPB1 supertypes		
	DP1	G11, K69, D84
	DP2	G11, E69, G84
	DP3	L11, K69, D84
	DP4	G11, K69, G84
	DP6	L11, E69, D84
	DP8	G11, E69, D84
DP serological categories		
	DP1	A56, E85, A86, V87
	DP2	E56, G85, P86, M87
	DP3	E56, E85, A86, V87
	DP4	A56, G85, P86, M87

Allele membership in a DPB1 supertype or DP serological category is defined by shared amino-acid sequences, whereas allele membership in a T-cell eptiope is defined by allele identity (name). Sequence definitions for DPB1 supertypes are derived from Taylor et al. (2009), defintions for DP serological categories are derived from Cano and Fernandez-Vina (2008), and T-cell epitope definitions are derived from Zino et al. (2007) and Sizzano et al. (2010).

a: Zino et al. (2007) and Sizzano et al. (2010) differ with respect to the assignment of DPB1*86:01 and *104:01; Zino et al. (2007) include these alleles in TCE3 group 2, while Sizzano et al. (2010) assign them to TCE3 group 3. Analyses using TCE3 epitopes as defined by Sizzano et al. (2010) are presented.

Table 2. Assignment of Observed DPB1 Alleles to Functional Categories

	T-Cell Epitope			
Allele	TCE3	TCE4	DPB1 Supertype[1]	DP Serological Category
DPB1*01:01	Group 3	Group 4	DP1	DP1
DPB1*02:01	Group 3	Group 3	DP2	DP2
DPB1*02:02	Group 3	Group 3	DP2	DP4
DPB1*03:01	Group 2	Group 2	DP3	DP3
DPB1*04:01	Group 3	Group 4	DP4	DP4
DPB1*04:02	Group 3	Group 4	DP4	DP2
DPB1*05:01	Group 3	Group 4	DP1	DP1
DPB1*06:01	Group 3	Group 4	DP6	DP3
DPB1*08:01	Group 3	Group 4	DP8	DP3
DPB1*09:01	Group 1	Group 1	DP6	DP3
DPB1*10:01	Group 1	Group 1	DP6	DP3
DPB1*11:01	Group 3	Group 4	11L:69R:84D	DP1
DPB1*13:01	Group 3	Group 4	DP6	DP1
DPB1*14:01	Group 2	Group 2	DP3	DP3
DPB1*15:01	Group 3	Group 4	11G:69R:84V	DP4
DPB1*16:01	Group 3	Group 4	DP8	DP3
DPB1*17:01	Group 1	Group 1	DP6	DP3
DPB1*18:01	Group 3	Group 4	11G:69K:84V	DP2
DPB1*19:01	Group 3	Group 4	DP8	DP1
DPB1*20:01	Group 3	Group 4	DP3	DP3
DPB1*21:01	Group 3	Group 4	DP6	DP1
DPB1*22:01	Group 3	Group 4	DP8	DP1
DPB1*23:01	Group 3	Group 4	DP4	DP4
DPB1*24:01	Group 3	Group 4	DP4	DP4
DPB1*25:01	Group 3	Group 4	DP3	DP3
DPB1*26:01	Group 3	Group 4	DP3	DP1
DPB1*27:01	Group 3	Group 4	DP3	DP1
DPB1*28:01	Group 3	Group 4	11G:69K:84V	DP2
DPB1*29:01	Group 3	Group 4	DP6	DP3
DPB1*30:01	Group 3	Group 4	DP6	DP1
DPB1*31:01	Group 3	Group 4	DP1	DP1
DPB1*32:01	Group 3	Group 4	DP2	DP2
DPB1*33:01	Group 3	Group 4	DP2	DP4
DPB1*34:01	Group 3	Group 4	11G:69K:84V	DP4

DPB1*35:01	Group 3	Group 4	DP3	DP3
DPB1*36:01	Group 3	Group 4	DP3	DP1
DPB1*37:01	Group 3	Group 4	DP6	DP3
DPB1*38:01	Group 3	Group 4	DP1	DP1
DPB1*39:01	Group 3	Group 4	DP4	DP4
DPB1*40:01	Group 3	Group 4	11G:69K:84V	DP4
DPB1*41:01	Group 3	Group 4	DP2	DP2
DPB1*44:01	Group 3	Group 4	DP6	DP3
DPB1*45:01	Group 2	Group 2	DP3	DP3
DPB1*46:01	Group 3	Group 4	DP2	DP2
DPB1*47:01	Group 3	Group 4	DP2	DP4
DPB1*48:01	Group 3	Group 4	DP2	DP2
DPB1*49:01	Group 3	Group 4	DP4	DP2
DPB1*50:01	Group 3	Group 4	DP1	DP3
DPB1*51:01	Group 3	Group 4	DP4	DP2
DPB1*52:01	Group 3	Group 4	DP3	DP1
DPB1*55:01	Group 3	Group 4	DP6	DP1
DPB1*56:01	Group 3	Group 4	DP3	DP1
DPB1*59:01	Group 3	Group 4	DP4	DP2
DPB1*60:01	Group 3	Group 4	DP4	DP2
DPB1*61:01N	None*	None*	None*	None*
DPB1*62:01	Group 3	Group 4	11G:69K:84V	DP4
DPB1*63:01	Group 3	Group 4	DP1	DP1
DPB1*65:01	Group 3	Group 4	DP1	DP1
DPB1*66:01	Group 3	Group 4	11L:69K:84G	DP4
DPB1*67:01	Group 3	Group 4	DP3	DP1
DPB1*68:01	Group 3	Group 4	DP1	DP3
DPB1*69:01	Group 3	Group 4	11L:69R:84D	DP3
DPB1*70:01	Group 3	Group 4	DP3	DP3
DPB1*72:01	Group 3	Group 4	DP4	DP4
DPB1*75:01	Group 3	Group 4	DP4	DP2
DPB1*76:01	Group 3	Group 4	DP3	DP3
DPB1*77:01	Group 3	Group 4	DP4	DP2
DPB1*79:01	Group 3	Group 4	DP3	DP3
DPB1*80:01	Group 3	Group 4	DP4	DP2
DPB1*81:01	Group 3	Group 4	DP2	DP2
DPB1*87:01	Group 3	Group 4	DP3	DP1
DPB1*88:01	Group 3	Group 4	DP6	DP3

DPB1*93:01	Group 3	Group 4	DP6	DP3
DPB1*111:01	Group 3	Group 4	DP3	DP3

*: The DPB1*61:01N allele name and peptide sequence was not included in these analyses. The nucleotide sequence of this allele is identical to that of DPB1*03:01:01 with the exception of a G to T nonsense (amber) mutation in codon 67, which results in a truncated protein product. DPB1*61:01N was reported in one individual in the Cameroon population.

1: The amino acid residues at positions 11, 69 and 84 are shown for those alleles that do not correspond to a DPB1 Supertype.

Table 3. Summary of Amino Acid-level Ewens-Watterson Analysis Based on *DPB1* Exon 2-encoded Peptide Sequences.

Amino acid Position	mean F_{nd}	Number of Variant Populations	Number of Populations with EW test p-values <0.05	mean k	Proportion of populations with $F_{nd} < 0$	p-value of parametric t-test	Significant Trend
8	-0.994	134	11	2	0.858	1.7E-27	-
9	-0.430	134	5	2.87	0.739	2.8E-08	-
11	-0.760	131	6	2	0.847	4.1E-22	-
12	0.369	1	0	2	0	N.D.	+
17	0.931	4	0	2	0	4.5E-06	+
32	0.915	1	0	2	0	N.D.	+
33	0.708	70	0	2	0	8.1E-37	+
35	-0.345	127	12	2.83	0.551	7.3E-05	-
36	-1.294	128	43	2	0.891	1.5E-34	-
55	-1.124	128	27	2.92	0.938	9.7E-38	-
56	-1.464	128	39	2	0.922	2.2E-47	-
57	-0.259	131	1	2.05	0.649	3.5E-05	-
65	-0.222	132	2	2.04	0.614	2.6E-04	-
69	-0.645	125	2	2.57	0.840	4.5E-17	-
72	0.789	16	0	2	0	3.9E-10	+
76	-0.301	133	1	2.76	0.684	1.6E-05.	-
84	-1.035	135	15	2.49	0.926	1.1E-37	-
85-87	-1.354	135	27	2	0.926	6.7E-50	-

Analytical results and summary statistics (described below) assessed for each of 18 polymorphic amino acid (AA) positions in a dataset of 136 populations are shown. These 18 AAs represent all of the DPB1 exon 2-encoded AA variation observed in the dataset. Invariant AA positions (displaying a single AA residue across all populations) are not shown. AA positions 85-87 are observed as a pair of invariant sequence blocks (G85-V86-M87 or E85-A86-V87), and are treated as a single polymorphic position.

Analytical Results and Summary Statistics:
mean F_{nd}: Average values of the normalized deviate of homozygosity (Fnd) for each AA position over the number of populations for which that AA position was polymorphic.

Number of Variant Populations: Describes the number of populations (out of 136) that display any polymorphism for a given position.
Number of Populations with EW test p-values < 0.05: Describes the number of populations (out of 136) for which any individual Ewens-Watterson (EW) homozygosity test displayed statistical significance (p-value < 0.05).
mean k: Describes the mean number of amino acid residues observed at a given position across populations for which that AA position was polymorphic.
Proportion of populations with $F_{nd} < 0$: Identifies the fraction of populations displaying homozygosity lower than the value expected under the EW model for a population of the same size, displaying the same number of alleles (polymorphic AAs) evolving under the null hypothesis of neutral evolution (Ho: $F_{nd} = 0$).
p-value of parametric t-test: Describes the p-value of a t-test comparing overall trends in in F_{nd} values with respect to the null hypothesis. For such parametric t-test comparisons of overall trends in F_{nd} between 476 locus-categories (DPB1 alleles, TCE3 and TCE4 T-cell epitopes, 18 individual AA positions, 91 AA pairs including DP SCs and 364 AA trios including *DPB1* Supertypes), significance was evaluated at the 1.05×10^{-4} level.
Significant Trend: Based on the significance levels of the t-tests, a trend toward positive, directional selection (+), negative, balancing selection (-), or neutral evolution (blank) is indicated.

N.D. Not determined, as the t-test cannot be calculated for single populations.

Table 4. Summarized Analyses of Selection for Functional Categories of DPB1 Alleles

Functional Category	TCE3 Groups	TCE4 Groups	Supertypes and 11-69-84 Triplets	DP Serologic Categories
Mean F_{nd}	0.171	-0.261	-0.495	-1.15
Mean p-value	0.535	0.438	0.349	0.164
Neutrality t-test p-value 1	0.006	9.0×10^{-5}	1.32×10^{-10}	2.93×10^{-39}
# F_{nd} < 0 [2]	41	92	109	124
# F_{nd} > 0 [2]	90	44	27	12
# p-value < 0.05 [2]	2	2	18	73
Mean k	2.58	3.46	6.18	3.80
# k = 1 [2]	5	0	0	0
# k = 2 [2]	48	19	9	9
# k = 3 [2]	83	35	7	9
# k = 4 [2]	--	82	9	118
# k = 5 [2]	--	--	17	--
# k = 6 [2]	--	--	26	--
# k = 7 [2]	--	--	26	--
# k = 8 [2]	--	--	36	--
# k = 9 [2]	--	--	6	--
# k = 10 [2]	--	--	0	--

1: The threshold of significance for neutrality t-tests is 0.05/539 (the number of EW tests of neutrality performed for populations with k>1) or 9.276×10^{-5}.

2: Out of a total of 136 populations.

-- : The maximum value of k is 3 for TCE3 Groups, 4 for TCE4 Groups and DP SCs, and 10 for DP STs.

Table 5. Fnd values for *DPB1* alleles and DP serologic categories

Region	Population	2n	DPB1 Alleles			DP Serologic Categories			Randomization Tests of Neutrality	
			k	F_{nd}	EW test p-value*	k	F_{nd}	EW test p-value*	mean k	fRT $F_{nd} \leq$ SC F_{nd}
SubSaharan Africa										
	Aka Pygmies	161	11	2.236	0.9616	4	0.150	0.5909	3.8192	0.6308
	CAR Aka Pygmies	186	11	2.040	0.9517	4	0.091	0.5711	3.8192	0.6282
	Cameroon	344	20	-0.556	0.3274	4	-1.470	**0.0496**	3.979	0.564
	Congo Kinshasa Bantu	179	18	-0.493	0.3615	4	-1.353	0.07	3.9768	0.5764
	Gabonese	240	21	0.169	0.6791	4	-1.304	0.0895	3.9894	0.5648
	Gambian	292	15	-0.226	0.5064	4	-1.539	**0.0315**	3.9396	0.2752
	Kenyan	246	34	1.132	0.8863	4	-1.716	**0.0076**	3.9988	0.2272
	Nigerian	260	23	3.462	0.9887	4	-1.014	0.173	3.9928	0.311
	Shona	456	20	-0.010	0.6094	4	-1.462	0.057	3.9778	0.4418
	Sudanese	193	13	-0.796	0.1981	4	-1.674	**0.0095**	3.8932	0.145
	Ugandan	94	14	-0.114	0.5679	4	-1.095	0.1395	3.9196	0.5832
	West African	200	10	-1.331	**0.0171**	4	-1.747	**0.0053**	3.7476	0.1818
	Zulu	174	14	-0.560	0.3272	4	-1.397	0.0554	3.9106	0.424
North Africa										
	Tunisia	200	18	0.022	0.6208	4	-1.834	**0.0018**	3.9716	**0.0126**
	Tunisian	202	17	-0.470	0.3770	4	-1.762	**0.0048**	3.9664	0.0788
Europe										
	Australian Caucasian[a]	100	15	0.680	0.8152	4	-1.593	**0.0118**	3.9368	**0.0114**
	Austria	980	19	0.193	0.6839	4	-1.784	**0.0089**	3.9822	**0.0444**
	Basques	195	15	-0.122	0.5567	4	-1.478	**0.0384**	3.9356	0.173
	Basques	192	37	2.789	0.9807	4	-1.606	**0.0151**	4	0.2524
	Belgium	197	16	0.070	0.6379	4	-1.693	**0.0083**	3.9518	**0.0272**

Breton	300	17	0.095	0.6512	4	-1.707	**0.0093**	3.9646	**0.0356**
British	374	19	0.806	0.8370	4	-1.698	**0.0108**	3.9806	**0.0052**
British	124	15	0.813	0.8356	4	-1.513	**0.0243**	3.938	**0.0408**
Catalan	169	17	0.658	0.8108	4	-1.620	**0.0128**	3.9648	**0.029**
Caucasian[a]	296	15	-0.403	0.4117	4	-1.775	**0.0057**	3.9388	0.0622
Caucasian[a]	475	17	0.057	0.6353	4	-1.770	**0.0074**	3.964	**0.0272**
Caucasian[a]	184	13	-0.376	0.4310	4	-1.709	**0.0069**	3.89	0.0522
CEPH[a]	266	17	0.220	0.6935	4	-1.699	**0.0094**	3.964	**0.0476**
CEPH[a]	248	18	0.458	0.7642	4	-1.677	**0.0109**	3.976	**0.0406**
Czech	204	18	0.826	0.8392	4	-1.553	**0.0227**	3.9616	0.1148
Danish	142	16	0.538	0.7828	4	-1.620	**0.0119**	3.9538	**0.021**
English	84	11	-0.273	0.4807	4	-1.647	**0.0074**	3.8084	**0.0244**
Finn	60	9	-0.429	0.4029	4	-1.496	**0.0185**	3.674	0.085
French	114	14	0.355	0.7357	4	-1.545	**0.0188**	3.9084	**0.038**
French	354	15	0.102	0.6523	4	-1.627	**0.018**	3.9408	**0.03**
German	411	17	0.016	0.6206	4	-1.810	**0.0053**	3.9672	**0.023**
German Essen	343	22	1.039	0.8742	4	-1.758	**0.0067**	3.992	**0.0272**
Greece	196	13	-0.153	0.5412	4	-1.283	0.0922	3.8898	0.3904
Greece	492	26	1.154	0.8915	4	-1.693	**0.0128**	3.9976	0.1822
Greek	95	14	0.499	0.7733	4	-1.522	**0.0199**	3.92	**0.0488**
Italian	174	15	-0.065	0.5814	4	-1.679	**0.0088**	3.9316	0.0572
Italy Central	759	20	-0.309	0.4671	4	-1.870	**0.005**	3.9836	0.1192
Italy North	100	14	0.901	0.8515	4	-1.401	**0.043**	3.9228	0.0662
North Ireland	300	16	1.012	0.8650	4	-1.459	0.0501	3.9562	**0.0138**
Norwegian	256	12	-0.251	0.4935	4	-1.663	**0.0114**	3.853	**0.038**
Sardinian	97	9	-0.716	0.2558	4	-1.382	**0.0466**	3.6764	0.3188

	Slavic	200	18	0.661	0.8058	4	-1.698	**0.0078**	3.9734	0.0616
	Slovak	292	15	0.381	0.7378	4	-1.477	**0.0464**	3.9346	0.1758
	Slovenian	200	16	1.143	0.8840	4	-1.413	0.0538	3.936	0.057
	Svans	160	16	-0.432	0.3978	4	-1.731	**0.005**	3.952	0.0888
	Swedish	347	15	0.242	0.6975	4	-1.609	**0.0197**	3.9386	0.071
	Swedish	400	19	0.623	0.8018	4	-1.699	**0.011**	3.9822	0.0564
	US Caucasian[a]	230	12	0.364	0.7264	4	-1.346	0.0762	3.8544	0.0772
	Whites US[a]	536	24	1.092	0.8833	4	-1.778	**0.0074**	3.9954	**0.0218**
Southwest Asia										
	East Indian	118	14	0.074	0.6404	4	-1.639	**0.0093**	3.9152	**0.0328**
	Saudi	98	9	2.000	0.9478	4	0.082	0.5854	3.675	0.5744
Southeast Asia										
	Chinese	168	12	-0.862	0.1684	4	-1.746	**0.0047**	3.8608	0.0956
	Chinese	80	11	-0.374	0.4371	4	-1.553	**0.0144**	3.81	0.0958
	Han Chinese	94	12	0.388	0.7372	4	-0.385	0.4373	3.8584	0.8038
	Jing Chinese	274	20	0.389	0.7443	4	-1.054	0.1607	3.9816	0.7132
	Lisu	222	19	-0.610	0.2900	4	-1.503	**0.0345**	3.9822	0.4846
	Malay	104	20	0.058	0.6356	4	-1.518	**0.0215**	3.9772	0.306
	Manchu	96	11	0.115	0.6527	4	-1.374	**0.0489**	3.8184	0.1364
	Miao Hmong	168	10	-0.025	0.5902	4	-0.746	0.2729	3.754	0.6058
	Naxi	192	19	-0.635	0.2742	4	-1.601	**0.0157**	3.9802	0.3512
	Nu	214	19	-0.612	0.2896	4	-1.492	**0.0364**	3.9822	0.4944
	NuChinese	144	12	-1.250	**0.0215**	4	-1.677	**0.0081**	3.8622	0.2316
	Pumi	102	17	3.078	0.9839	3	-0.918	0.2462	3.9576	0.2394
	Shandong Han Chinese	196	17	0.223	0.6951	4	-1.624	**0.0131**	3.9604	0.1066

	Shanghai Chinese	206	11	-0.905	0.1521	4	-1.684	**0.0093**	3.816	0.1416
	Taiwanese	96	7	1.044	0.8456	3	0.058	0.5285	3.4444	0.5814
	Yao	132	10	0.364	0.7193	4	-0.439	0.4202	3.7496	0.7054
Oceania										
	Borneo	42	10	0.510	0.7720	4	-0.406	0.4267	3.7606	0.7208
	Cook Islands	100	8	-0.175	0.5277	4	-1.326	0.066	3.5882	0.1786
	East Timorese	172	9	-0.619	0.3064	4	-0.423	0.4236	3.6878	0.856
	Filipino	188	14	0.694	0.8126	4	0.076	0.5668	3.8874	0.8602
	Filipino	180	12	-0.248	0.4967	4	-0.195	0.4925	3.87	0.8952
	Indonesia	264	12	-0.801	0.2049	4	-0.518	0.3873	3.864	0.9334
	Javanese	118	16	-0.677	0.2530	4	-1.502	**0.0248**	3.9526	0.3608
	Maori	398	19	1.087	0.8781	4	-1.623	**0.0193**	3.979	0.1094
	Mixed Hawaiian	76	11	-0.510	0.3636	4	-1.260	0.0825	3.8256	0.3608
	Moluccan	92	12	-0.416	0.4122	4	-0.207	0.5006	3.818	0.9362
	PNG Highland	56	10	0.249	0.6936	4	-1.441	**0.0261**	3.7476	0.0728
	PNG Highland	176	6	-1.085	0.1115	4	-1.624	**0.0125**	3.262	0.2221
	PNG Lowland	96	10	-0.215	0.5068	4	-1.462	**0.0286**	3.7636	0.149
	Roro	52	6	1.480	0.9054	4	0.569	0.7178	3.2736	0.6937
	Samoa	100	6	-0.231	0.4956	4	-0.970	0.1833	3.2588	0.2999
	Samoan	58	6	0.477	0.7377	4	-0.357	0.4522	3.27	0.5298
	Tokelau	100	6	-0.734	0.2649	4	-1.391	**0.0448**	3.2626	0.2535
	Tolai	96	5	-0.538	0.3526	3	-1.430	0.0654	3.0262	0.1173
	Tonga	100	7	-0.103	0.5539	4	-1.036	0.1589	3.443	0.3840
	Trobriand Islanders	162	4	1.934	0.9935	3	1.429	0.9674	2.7262	0.8156
	Western Samoa	44	6	1.535	0.9057	4	0.204	0.643	3.2882	0.4778
Australia										

	Cape York	192	10	-0.269	0.4783	4	-0.552	0.3735	3.7534	0.7802
	Kimberley	76	5	0.166	0.6446	3	-0.164	0.4788	3.0254	0.5921
Northeast Asia										
	Japanese	420	14	-0.164	0.5413	4	-1.653	**0.0157**	3.9176	0.0958
	Japanese	100	10	-0.544	0.3445	4	-1.448	**0.0317**	3.7606	0.223
	Japanese	102	13	-0.664	0.2687	4	-1.725	**0.0035**	3.8906	0.0652
	Japan Fukuoka	172	12	0.123	0.6494	4	-1.441	**0.0445**	3.8602	0.12
	Khalkh Mongolian	81	9	-0.984	0.1178	4	-1.484	**0.0237**	3.6718	0.2748
	Khoton Mongolian	164	10	-0.648	0.2890	4	-1.664	**0.0096**	3.7504	0.1326
	South Korea	648	13	-0.704	0.2541	4	-1.706	**0.0135**	3.8916	0.2026
	South Korea	414	13	-0.661	0.2749	4	-1.665	**0.0142**	3.892	0.2002
North America										
	Canoncito	80	6	-0.495	0.3752	4	-0.477	0.3945	3.2584	0.7109
	Maya	30	5	1.478	0.9099	4	0.728	0.78	3.0414	0.6981
	Mixe	104	5	2.036	0.9572	3	1.377	0.9161	2.7154	0.9062
	Mixteco	104	8	2.376	0.9655	4	0.958	0.7816	3.4374	0.8445
	Pima	34	4	-0.815	0.2338	2	-1.565	0.1016	2.2856	0.4128
	Pima	190	9	0.115	0.6369	4	-0.869	0.2218	3.6752	0.5482
	Sioux	164	10	-0.604	0.3142	4	-1.454	**0.0397**	3.6754	0.281
	Zapotec	144	11	2.702	0.9788	4	-0.047	0.5392	3.7534	0.4588
	Zuni	100	4	-0.472	0.4032	4	-0.472	0.4032	2.72	0.7532
South America										
	Chiriguanos	108	12	1.118	0.8801	4	-1.132	0.1286	3.8604	0.0782

	Arsario	100	2	-1.222	0.1986	2	-1.222	0.1986	1.7444	1.0
	Bari	196	7	-1.037	0.1227	4	-1.318	0.0819	3.4434	0.4178
	Cayapa	166	7	-0.132	0.5390	4	-0.895	0.212	3.4366	0.5305
	Central-America	110	7	0.888	0.8215	4	-0.266	0.4789	3.4416	0.3640
	Coreguaje	90	2	-1.832	**0.0202**	2	-1.832	**0.0202**	1.7444	1.0
	Eastern Toba	270	12	0.981	0.8545	4	-0.625	0.348	3.8616	0.4518
	Embera	98	3	-1.356	0.0800	3	-1.356	0.08	2.3076	0.5970
	Ijka	80	2	-1.749	0.0675	2	-1.749	0.0675	1.7444	1.0
	Kogui	100	2	-1.621	0.1145	2	-1.621	0.1145	1.7444	1.0
	Mataco	120	5	0.111	0.6199	3	-0.438	0.4002	3.0388	0.3819
	Mataco Wichi	94	7	0.575	0.7633	4	-0.026	0.5558	3.4464	0.6619
	Ticuna	98	8	-0.665	0.2902	4	-1.143	0.1244	3.573	0.442
	Vaupes	92	2	-0.002	0.4202	2	-0.002	0.4202	1.7444	1.0
	Warao	68	3	0.497	0.6388	2	-0.278	0.3734	2.2974	0.2070
	Wayuu	108	3	-0.992	0.2082	3	-0.992	0.2082	2.3036	0.8004
	Xavantes	141	5	1.169	0.8395	4	0.659	0.7123	3.0276	0.5864
	Yucpa	146	2	-0.935	0.2376	2	-0.935	0.2376	1.7444	1.0
	Yucpa	232	2	-0.788	0.2531	2	-0.788	0.2531	1.7444	1.0
Other										
	African American[a]	481	24	-0.108	0.5600	4	-1.717	**0.0104**	3.9974	0.2624
	Blacks US[a]	385	24	-0.323	0.4522	4	-1.748	**0.0075**	3.9968	0.2592
	Ecuadorian Africans[a]	116	12	0.563	0.7822	4	-0.804	0.2445	3.8496	0.5256
	Martinique[a]	200	19	-0.412	0.4129	4	-1.793	**0.0035**	3.9824	0.0704
	MexicanAmerican[a]	225	18	0.595	0.7923	4	-1.470	**0.0424**	3.9736	0.1886

For each of 136 populations, the number of alleles (k), normalized deviate of homozygosity (F_{nd}) and p-values of the Ewens-Watterson (EW) homozygosity test are presented for DPB1 alleles and DP Serologic Categories (SCs), along with the results of a randomization test of neutrality (fRT F_{nd} ≤ SC F_{nd} and associated mean k values).

a: Migrant Population, as discussed in section 3.2.
*: P-values of the Ewens-Watterson homozygosity test of neutrality shown in bold are significant at the 0.05 level; no p-values are significant when corrected for multiple (136) comparisons (0.00036 level); however, as noted in section 2.2.7, such corrections are overly conservative, as these tests are not independent.

fRT F_{nd} ≤ SC F_{nd}: For each population, this value identifies the fraction of F_{nd} values generated under the randomization test (RT F_{nd}) that are less than or equal to the observed F_{nd} values for DP Serologic Categories (SC F_{nd}). Populations for which the SC F_{nd} values are lower than a significant fraction (< 0.05) of RT F_{nd} values (excluding those RT F_{nd} values for which k =1) are indicated in bold.

Table 6. Correlation Between DP Serologic Category Frequencies

Region		DP1	DP2	DP3	DP4
All Populations[a] (n = 123)	**DP1 SC**		**1.7×10^{-13}**	**2.55×10^{-07}**	**4.68×10^{-05}**
	DP2 SC	-0.60		0.72	**6.72×10^{-05}**
	DP3 SC	-0.44	-0.03		0.48
	DP4 SC	-0.36	-0.35	-0.06	
SubSaharan Africa (n = 13)	**DP1 SC**		0.0034	0.51	0.0043
	DP2 SC	-0.75		0.0024	0.28
	DP3 SC	0.20	-0.76		0.92
	DP4 SC	-0.73	0.32	-0.03	
Europe (n = 31)	**DP1 SC**		**5.90×10^{-11}**	0.10	0.0011
	DP2 SC	-0.88		0.40	**1.02×10^{-05}**
	DP3 SC	-0.30	0.16		**5.76×10^{-06}**
	DP4 SC	0.56	-0.70	-0.72	
Southeast Asia (n = 16)	**DP1 SC**		**2.90×10^{-05}**	0.34	0.0026
	DP2 SC	-0.85		0.84	0.30
	DP3 SC	-0.26	-0.05		0.51
	DP4 SC	-0.70	0.28	0.18	
Oceania (n = 21)	**DP1 SC**		**1.02×10^{-05}**	0.20	**7.07×10^{-08}**
	DP2 SC	-0.81		0.47	0.03
	DP3 SC	-0.29	-0.17		0.10
	DP4 SC	-0.89	0.49	0.37	
Northeast Asia (n = 8)	**DP1 SC**		0.04	0.80	0.02
	DP2 SC	-0.74		0.61	0.46
	DP3 SC	-0.11	0.22		0.33
	DP4 SC	-0.81	0.31	-0.40	
North America (n = 9)	**DP1 SC**		0.38	0.35	0.94
	DP2 SC	-0.33		0.40	0.0024
	DP3 SC	0.36	-0.32		0.66
	DP4 SC	0.03	-0.87	-0.17	
South America (n = 19)	**DP1 SC**		0.62	0.02	0.94

DP2 SC	0.12		**1.95×10^{-06}**	0.33
DP3 SC	-0.55	-0.86		0.84
DP4 SC	0.02	-0.24	-0.05	

For each regional matrix, correlations (r) are shown in the lower half, while the significances of these correlations (p-values) are shown in the upper half. Significance is assessed at the 7.58×10^{-04} level after correcting for 66 comparisons. Significant p-values are indicated in bold. Correlations for the North Africa, Southwest Asia, and Australia regions are not shown, as each of these regions comprised only two populations

a: Migrant populations, as defined in Table 3 and discussed in section 3.2, have been excluded from these comparisons.

Table 7. *DPA1*~DP Serologic Category Haplotypes and LD in Ten Populations From Five World Regions

DP SC~DPA1 Haplotype				Sub Saharan Africa	Europe	Oceania					North America		South America
DP SC	*DPB1* AA Position 85-87 Sequence	*DPA1* Allele	*DPA1* AA 31-50-83 Sequence	Kenya (2n 244)	Slovenia (2n 200)	East Timor (2n 172)	Filipino (2n 188)	Moluccas (2n 92)	PNG Highland (2n 156)	PNG Lowland (2n 96)	Pima (2n 30)	Pima (2n 190)	Ticuna (2n 92)
DP1	EAV	*01:03*	M-Q-T	0.015	0.031	0.006	0.011		0.006			0.005	
		02:01	Q-R-A	**0.133**	0.054	**0.194**	0.054	0.065	0.019	0.063		0.011	0.033
		02:02	Q-R-A	**0.165**	0.005	**0.478**	**0.690**	**0.630**	**0.160**	**0.344**		0.016	
		03:01	M-Q-T	0.003									
		04:01	M-R-A	0.004		0.025	0.021	0.011					
DP2	GPM	*01:03*	M-Q-T	**0.145**	**0.260**	**0.105**	0.048	**0.109**	**0.359**	**0.146**	**0.600**	**0.589**	**0.511**
		02:01	Q-R-A	0.005									
		02:02	Q-R-A	0.018									
		03:01	M-Q-T	**0.201**									
		03:02	M-Q-T	0.004									
		04:01	M-R-A	0.012			0.006	0.011					
DP3	EAV	*01:03*	M-Q-T	0.096	0.095	0.017	0.042	0.011	0.038	0.073		0.053	0.033
		02:01	Q-R-A	0.063	0.050	0.015	0.042					0.021	**0.174**
		02:02	Q-R-A	0.019	0.005	0.033	0.012	0.033		0.010			0.076
		02:03	M-R-A	0.004									
		03:01	M-Q-T	0.016									
		04:01	M-R-A			0.016	0.005	0.021	0.045	0.021			
DP4	GPM	*01:03*	M-Q-T	0.089	**0.484**	**0.104**	0.058	**0.109**	**0.372**	**0.344**	**0.400**	**0.305**	**0.174**
		01:04	M-Q-T		0.005		0.005						
		02:01	Q-R-A	0.004	0.011								
		02:02	Q-R-A			0.006	0.005						
		03:01	M-Q-T	0.005									
	Diversity			0.87	0.68	0.71	0.51	0.57	0.70	0.73	0.48	0.56	0.67

Estimated haplotype frequencies for DPA1~DP Serologic Category (SC) haplotypes are shown for 10 populations from 5 world regions. The table is divided into four large rows, which represent the four DP SCs (DP1 to DP4, in the first column). The amino acid (AA) sequence of DPβ positions 85-87 is shown in the second column. The *DPA1* alleles observed in haplotypes with each DP SC are shown in the third column. The AA sequences of DPα positions 31, 50 and 83 are shown in the fourth column. *DPA1*~DP SC haplotypes that were not observed in any populations are not shown. Blank cells indicate haplotypes that were not observed in specific populations (frequency = 0). Haplotypes with frequencies > 0.1 are shown in bold. Dark grey cells indicate low haplotype-level linkage disequilibrium (LD) values ($D'_{ij} < 0.5$), while light grey cells indicate high LD values ($D'_{ij} > 0.5$). Haplotypes with frequencies > 0.1 are shown in bold. Haplotype diversity measures are shown in the bottom row for each population. Diversity is represented as heterozygosity, or $1 - \Sigma p_i^2$ where p_i is each haplotype frequency in a population.

Table 8. Summary of Evidence for Selection Based on EW tests and F_{nd} Values

Region (# pops)	DPB1 Alleles	DP Serologic Categories	TCE3 T-Cell Epitopes	TCE4 T-Cell Epitopes	DPB1 Supertypes
All	Weak	Strong	Weak	Weak	Weak
	Fnds mostly positive	Fnds mostly negative	Fnds mostly positive	Fnds slightly negative	Fnds somewhat negative
(136)	2% pops significant	54% pops significant	2% pops significant	2% pops significant	13% pops significant
SSA	Weak	Moderate	Weak	Weak	Weak
	Fnds mostly positive	Fnds mostly negative	Fnds mostly positive	Fnds slightly negative	Fnds somewhat negative
(13)	8% pops significant	38% pops significant	0% pops significant	0% pops significant	8% pops significant
EUR	Weak	**Very Strong**	Weak	Weak	Weak
	Fnds mostly positive	**Fnds quite negative**	Fnds mostly positive	Fnds somewhat negative	Fnds somewhat negative
(39)	0% pops significant	**90% pops significant**	0% pops significant	0% pops significant	0% pops significant
SEA	Weak	Strong	Weak	Weak	Strong
	Fnds mostly positive	Fnds quite negative	Fnds mostly positive	Fnds somewhat negative	Fnds mostly negative
(16)	6% pops significant	63% pops significant	0% pops significant	0% pops significant	56% pops significant
OCE	Weak	Slight	Weak	Weak	Weak
	Fnds mostly positive	Fnds somewhat negative	Fnds mostly positive	Fnds mostly positive	Fnds somewhat negative
(21)	0% pops significant	29% pops significant	0% pops significant	0% pops significant	10% pops significant
NEA	Weak	**Very Strong**	Weak	Weak	Slight
	Fnds somewhat negative	**Fnds quite negative**	Fnds mostly positive	Fnds somewhat negative	Fnds somewhat negative
(8)	0% pops significant	**100% pops significant**	0% pops significant	0% pops significant	25% pops significant
NAM	Weak	Weak	Weak	Weak	Weak
	Fnds mostly positive	Fnds somewhat negative	Fnds mostly positive	Fnds mostly positive	Fnds mostly positive
(9)	0% pops significant	11% pops significant	0% pops significant	0% pops significant	0% pops significant
SAM	Weak	Weak	Weak	Weak	Weak
	Fnds somewhat negative	Fnds somewhat negative	Fnds somewhat negative	Fnds somewhat negative	Fnds somewhat negative
(19)	5% pops significant	5% pops significant	11% pops significant	11% pops significant	5% pops significant

The evidence of balancing selection for each investigated category of variation (individual alleles, SCs, TCE3s, TCE4s, and STs) is summarized for all populations, as well as for subsets of populations in each of seven world regions, based on the proportion of positive or negative Fnd values and the percentage of populations displaying significant Fnd values, indicating deviation from the expectation of neutrality. Regions with fewer than six populations were not included. Evidence of selection is categorized as weak, slight, moderate, strong, and very strong based on the percentage of populations displaying significant Fnd values. Cases where evidence of balancing selection is greater than “weak” (negative Fnd and >24% populations displaying significance) are indicated with grey shading ranging from light to dark.