Ancient Trans-Species Polymorphism at the Major Histocompatibility Complex in Primates

Alyssa Lyn Fortier[1,2*] **and Jonathan K. Pritchard**[1,2]

***For correspondence:**
afortier@stanford.edu (ALF);
pritch@stanford.edu (JKP)

[1]Department of Biology, Stanford University, Stanford, CA USA; [2]Department of Genetics, Stanford University, Stanford, CA USA

Abstract

Classical genes within the Major Histocompatibility Complex (MHC) are responsible for peptide presentation to T cells, thus playing a central role in immune defense against pathogens. These genes are subject to strong selective pressures including both balancing and directional selection, resulting in exceptional genetic diversity—thousands of alleles per gene in humans. Moreover, some alleles appear to be shared between primate species, a phenomenon known as trans-species polymorphism (TSP) or incomplete lineage sorting, which is rare in the genome overall. However, despite the clinical and evolutionary importance of MHC diversity, we currently lack a full picture of primate MHC evolution. To start addressing this gap, we explore variation *across* genes and species in our companion paper (***Fortier and Pritchard, 2024***) and here we explore variation *within* individual genes. We used Bayesian phylogenetic methods to determine the extent of TSP at 17 MHC genes, including classical and non-classical Class I and Class II genes. We find strong support for deep TSP in 7 of 10 classical genes, including—remarkably—between humans and old-world monkeys in MHC-DQB1. Despite the long-term persistence of ancient lineages, we additionally observe rapid evolution at nucleotides encoding the proteins' peptide-binding domains. The most rapidly-evolving amino acid positions are extremely enriched for autoimmune and infectious disease associations. Together, these results suggest complex selective forces—arising from differential peptide binding—that drive short-term allelic turnover within lineages while also maintaining deeply divergent lineages for at least 31 million years.

Introduction

The Major Histocompatibility Complex (MHC) is a large locus containing many immune genes that is shared among the jawed vertebrates (***Radwan et al., 2020***). In humans, the MHC is also known as the HLA (Human Leukocyte Antigen) region; it spans about 5 megabases (Mb) on chromosome 6 and contains 412 genes (***Figure 1***A) (***MHC, 2022***; ***O'Leary et al., 2016***). Many of these are part of the MHC gene family, a large group of evolutionarily-related genes with varying functions. The "classical" MHC genes are responsible for presenting protein fragments for inspection by T cells. MHC peptide presentation allows T cells to monitor the body for the presence of foreign peptides, which might indicate infection or cancer; this is crucial for vertebrate immune surveillance (***Neefjes et al., 2011***). "Non-classical" MHC genes are essential to the innate immune system, where they perform a variety of niche roles.

The MHC locus is extraordinarily polymorphic. Haplotypes can vary widely in gene content, and thousands of distinct alleles are observed at the classical genes in humans and other primates

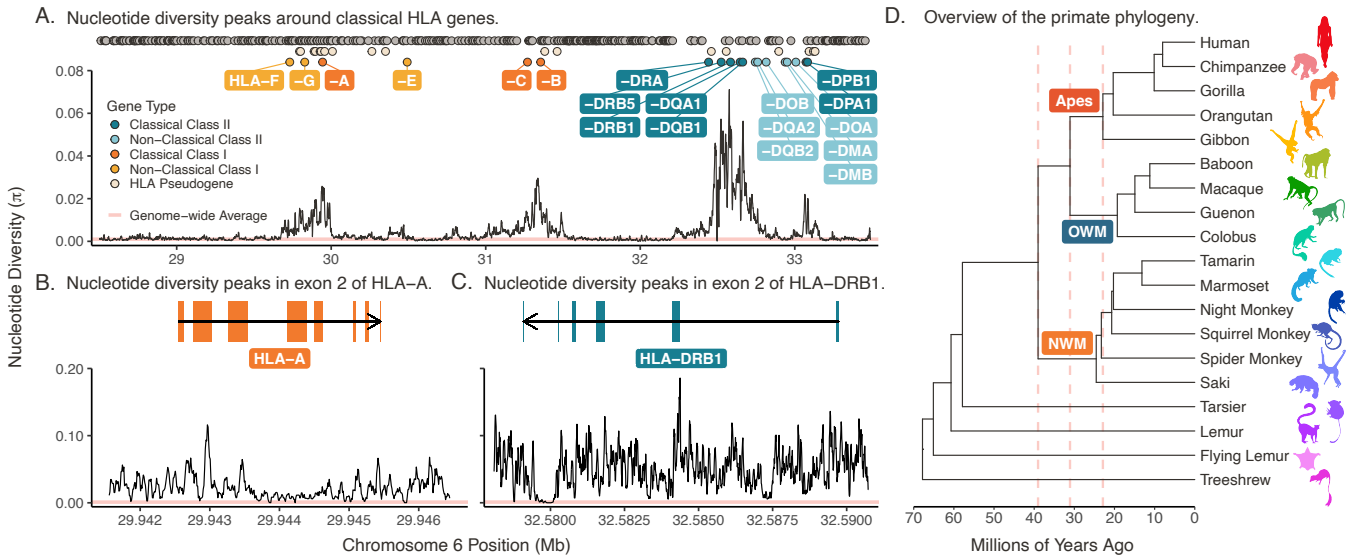

Figure 1. The MHC region in humans (HLA). A) Each point at top represents the location of a gene. The different types of HLA genes are distinguished by different colors, shown in the key at left. The 19 functional HLA genes are labeled with their name (omitting their "HLA" prefix due to space constraints). Gray points represent non-HLA genes and pseudogenes in the region. The black line shows nucleotide diversity (Nei and Li's π) across the region, while the pink horizontal line shows the genome-wide average nucleotide diversity ($\pi \approx 0.001$) (***Sachidanandam et al., 2001***). **B)** Nucleotide diversity around classical Class I gene HLA-A, with exon structure shown. **C)** Nucleotide diversity around classical Class II gene HLA-DRB1, with exon structure shown. **D)** Species tree showing the phylogenetic relationships among selected primates from this study (***Kuderna et al., 2023***). The colors of the icons are consistent with colors used throughout the paper to distinguish species. The pink vertical dashed lines indicate split times of the NWM from the apes/OWM (39 MYA), OWM from the apes (31 MYA), and the lesser apes (gibbons) from the great apes (23 MYA).

Figure 1—figure supplement 1. Nucleotide diversity around the rest of the genes.

Figure 1—figure supplement 2. Color and abbreviation key/tree for all species used in this study.

(***Maccari et al., 2017***, ***2020***; ***Robinson et al., 2019***). Different alleles are functionally diverse, with distinct peptide-binding affinities and, consequently, allelic differences in pathogen detection (***Neefjes et al., 2011***; ***Adams and Luoma, 2013***). Given this huge diversity of functionally distinct alleles, the MHC is by far the most important locus in the genome for inter-individual variation in both infectious and autoimmune disease risk, with thousands of GWAS hits (***Buniello et al., 2019***; ***Smith et al., 2024***). In our companion paper (***Fortier and Pritchard, 2024***), we explored the relationships between the different classical and non-classical genes. Here, we look within 17 specific genes to characterize trans-species polymorphism, a phenomemon characteristic of long-term balancing selection.

Historically, the MHC provided some of the first clear examples of positive selection in early studies of molecular evolution. By the 1980s and 1990s, researchers had noted an excess of missense variants (i.e., $dN/dS > 1$) in the peptide-binding regions of classical MHC genes (***Hughes and Nei, 1988***, ***1989***), alleles shared across species (***Arden and Klein, 1982***; ***Mayer et al., 1988***), and high nucleotide diversity across the region (***Wakeland et al., 1987***; ***Nei and Hughes, 1991***) in rodents and primates. Indeed, modern data show that nucleotide diversity in the human MHC (HLA region) exceeds 70-times the genome-wide average near the classical genes, suggesting ancient balancing selection (***Figure 1***A-C). Meanwhile, the MHC also features prominently in genome-wide scans for short-term directional selection (***Mathieson et al., 2015***; ***Field et al., 2016***; ***Allentoft et al., 2022***; ***Cong et al., 2022***; ***Okada et al., 2018***; ***Yasumizu et al., 2020***).

In the present paper, we explore a particularly striking feature of the selection signals at MHC, namely the evidence for extremely deep coalescence structure. Some alleles (haplotypes) are more closely related to corresponding alleles from another species than they are to distinct alleles from their own species. This phenomenon is referred to as *trans-species polymorphism* (TSP).

TSP is rare overall in humans. Across most of the genome, human alleles coalesce to a common ancestor well within the human lineage, typically around 2 million years ago (***Mallick et al., 2016***). Indeed, only ∼100 loci genome-wide show compelling evidence for sharing of ancestral alleles between humans and our closest relatives, chimpanzees (***Leffler et al., 2013***). TSP among humans and more distantly-related species is even rarer; besides the MHC, the only other clear example of deep TSP is at the ABO locus (which influences blood type) (***Azevedo et al., 2015***). At this locus, both the A and B alleles are shared by descent throughout the apes, implying that the A and B lineages date back to at least the divergence-point of humans and gibbons 23 million years ago (***Ségurel et al., 2012***; ***Kuderna et al., 2023***). Such deep coalescence is extraordinarily unlikely under a neutral model, and instead points to some form of balancing selection.

Meanwhile, TSP is evident at multiple MHC genes, and in many different phylogenetic clades. TSP at this locus was first proposed in the 1980s on the basis of unusual sequence similarity between mice and rats (***Klein, 1980***; ***Arden and Klein, 1982***; ***Klein, 1987***; ***Figueroa et al., 1988***; ***McConnell et al., 1988***; ***Wakeland et al., 1987***), and between humans and chimpanzees (***Lawlor et al., 1988***; ***Mayer et al., 1988***). Later work has reported likely TSP between humans and apes (***Slierendregt et al., 1995***; ***Boyson et al., 1996***; ***McKenzie et al., 1999***; ***Wroblewski et al., 2017***) and humans and old world monkeys (***Mayer et al., 1992***; ***Brändle et al., 1992***; ***Kupfermann et al., 1992***; ***Slierendregt et al., 1992***; ***Geluk et al., 1993***; ***Satta et al., 1996***; ***Otting et al., 2000***; ***Kriener et al., 2000***; ***Gyllensten et al., 1990***; ***Kriener et al., 2001***; ***Otting et al., 1992***; ***Otting and Bontrop, 1995***; ***Otting et al., 1992***, ***2002***); deep TSP is also consistent with the high levels of genetic diversity within the MHC. Such ancient TSP would make the MHC unique compared to any other locus in the genome. However, most previous work has not fully accounted for the inherent uncertainty in phylogenetic inference, especially given the potential for convergent evolution at functional sites. Although there is clear evidence for TSP, its exact age at each gene is still uncertain.

To address these questions, we used data from the IPD-MHC/HLA database—a large repository for MHC allele sequences from humans, non-human primates, and other vertebrates—along with supplementary sequences from NCBI RefSeq (***Maccari et al., 2017***, ***2020***; ***Robinson et al., 2019***). This represents the most complete sampling of primate MHC genes to date, spanning the entire primate tree (***Figure 1***D). We account for the uncertainty in phylogenetic inference using a Bayesian MCMC approach (*BEAST2*) which is well-suited to handle highly variable and rapidly-evolving sequences. In our companion paper (***Fortier and Pritchard, 2024***), we compared genes across dozens of species and learned more about the orthologous relationships among them. That process helped inform this work, where we assess support for TSP within individual genes.

We find support for TSP among the African apes for genes MHC-C, -DPA1, and -DRB3, among the great apes for MHC-DPB1, and among all apes for MHC-B. We also find conclusive evidence for TSP at least back to the ancestor of humans and OWM in MHC-DQB1, implying—remarkably—that alleles have been maintained by balancing selection for at least 31 million years. Rapidly-evolving sites are mainly located in the critical peptide-binding regions of the classical genes, but are spread throughout the coding region of the non-classical genes. Moreover, the most rapidly-evolving sites are also frequently associated with immune phenotypes and diseases in the literature, connecting our evolutionary findings with their functional consequences. These results highlight the contrasting roles of ancient balancing selection and short-term directional selection within the peptide-binding regions of the classical genes, and motivate further evolutionary and functional studies to better understand this unique system.

Results

Data

We collected MHC nucleotide sequences for all genes from the IPD-MHC/HLA database, a large repository for MHC alleles from humans, non-human primates, and other vertebrates (***Maccari et al., 2017***, ***2020***; ***Robinson et al., 2024***). Although extensive, this database includes few or no se-

quences from important primates such as the gibbon, tarsier, and lemur. Thus, we supplemented our set of alleles using sequences from NCBI RefSeq (***O'Leary et al., 2016***). Because the MHC genes make up an evolutionarily-related family, they can all be aligned (***Kaufman, 2022***; ***Adams and Luoma, 2013***). In our companion paper (***Fortier and Pritchard, 2024***), we utilized these large multigene alignments for Class I, Class IIA, and Class IIB to compare genes. Here, we analyze subsets of those alignments, each focusing on a gene or group of closely-related genes.

We considered 16 gene groups and two or three different genic regions for each group: exon 2 alone, exon 3 alone, and/or exon 4 alone. Exons 2 and 3 encode the peptide-binding region (PBR) for the Class I proteins, and exon 2 alone encodes the PBR for the Class II proteins. For the Class I genes, we also considered exon 4 alone because it is comparable in size to exons 2 and 3 and provides a good contrast to the PBR-encoding exons. See the Methods for more detail on how gene groups were defined. Because few intron sequences were available for non-human species, we did not include them in our analyses.

Trans-Species Polymorphism is Widespread.

For each gene group and genic region, we used the Bayesian phylogenetics software *BEAST2* (***Bouckaert et al., 2014***, ***2019***) with package *SubstBMA* (***Wu et al., 2013***) to infer phylogenies. One major advantage of *BEAST2* over less tuneable methods is that it can allow evolutionary rates to vary across sites, which is important for genes such as these which experience rapid evolution in functional regions (***Wu et al., 2013***). We also considered each exon separately to minimize the impact of recombination as well as to compare and contrast the binding-site-encoding exons with non-binding-site-encoding exons.

We can visualize each set of phylogenies as a single summary tree, which maximizes the sum of posterior clade probabilities (***BEA, 2024***). Three of these summary trees are shown in ***Figure 2***, constructed from the second exons of classical Class I gene MHC-C, classical Class II gene MHC-DQB, and non-classical Class II gene MHC-DOA, respectively (see ***Figure 2—figure Supplement 1***-***Figure 2—figure Supplement 10*** for the other exons and genes).

In ***Figure 2***, each tip represents a sequence, labeled with its allele name (see ***Appendix 1*** for more details on nomenclature) or its RefSeq accession number. Groups of tips containing similar alleles are collapsed for clarity. Preceding the allele label is a colored rectangle with a four-letter abbreviation indicating the species; the species key is in the middle of the plot and also displays the known species tree (***Kuderna et al., 2023***).

MHC-C is a classical Class I gene which duplicated from MHC-B in the ancestor of the great apes (***Piontkivska and Nei, 2003***; ***Fukami-Kobayashi et al., 2005***; ***Abi-Rached et al., 2010***; ***Adams and Parham, 2001***; ***Lugo and Cadavid, 2015***). Its protein product participates in classical antigen presentation and also serves as the dominant Class I molecule for interacting with killer cell immunoglobulin-like receptors (KIRs) in innate immunity (***Adams and Parham, 2001***; ***Guethlein et al., 2015***; ***Vollmers et al., 2021***). MHC-DQB is a classical Class II gene which pairs with MHC-DQA. Apes have two MHC-DQ copies, MHC-DQA1/MHC-DQB1 and MHC-DQA2/MHC-DQB2, while the second copy was deleted in OWM. NWM can have two or three sets of MHC-DQ genes, depending on species, but it has been unclear whether any of them are 1:1 orthologous with the ape or OWM genes. Lastly, MHC-DOA is a non-classical Class II gene whose protein product modulates MHC-DM activity, indirectly affecting Class II peptide presentation (***Heijmans et al., 2020***; ***Neefjes et al., 2011***). The genes' differing roles result in different patterns in the phylogenetic trees.

Critically, we observe that, at classical genes MHC-C (***Figure 2***A) and MHC-DQB (***Figure 2***B), the alleles fail to cluster together according to species, as indicated by the mixed-color clades throughout the trees. In MHC-C, human (HLA; red), chimpanzee (Patr; dark pink), bonobo (Papa; light pink), and even gorilla (Gogo; orange) alleles can be found throughout the tree, indicating that variation in this gene is almost as old as the gene itself.

Figure 2B displays the *BEAST2* tree consisting of MHC-DQB1, -DQB2, and outgroup -DQB alleles all together. It shows many mixed-color clades throughout, consisting of ape (red/orange/yellow)

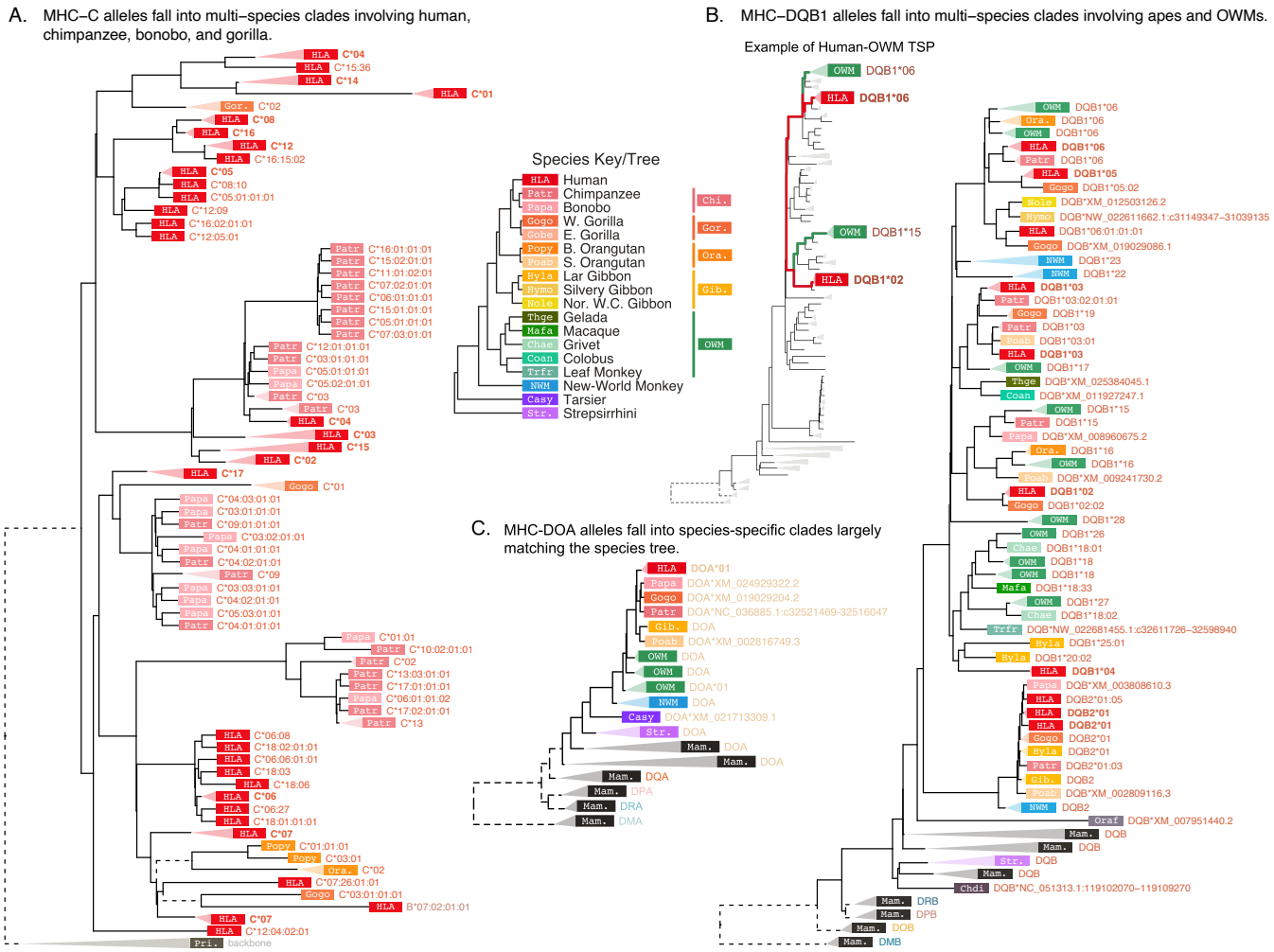

Figure 2. *BEAST2* allele summary trees using sequences from exon 2. A) MHC-C, **B)** MHC-DQB1, and **C)** MHC-DOA. Each tip represents an allele, with color and four-letter abbreviation representing the species. The species label is followed by the allele name. For simplicity, monophyletic groups of similar alleles are collapsed with a triangle and labeled with their one-field allele name. The color/abbreviation key (center) also depicts the species tree (***Kuderna et al., 2023***). Human alleles (HLA; red) are bolded for emphasis. Dashed outgroup branches are scaled by a factor of $\frac{1}{10}$ to clarify tree structure within the clade of interest. The smaller inset tree in panel B highlights the relationships between two human allele groups (red) and two OWM allele groups (green). The indicated human and OWM lineages coalesce more recently between groups than within each group. Pri., primate backbone sequences; Mam., mammal outgroup sequences.

Figure 2—figure supplement 1. MHC-A-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 2. MHC-B-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 3. MHC-C-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 4. MHC-E-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 5. MHC-F-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 6. MHC-G-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 7. MHC-DRA-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 8. MHC-DQA-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 9. MHC-DPA-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 10. MHC-DMA-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 11. MHC-DOA-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 12. MHC-DRB-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 13. MHC-DQB-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 14. MHC-DPB-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 15. MHC-DMB-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

Figure 2—figure supplement 16. MHC-DOB-Related *BEAST2* tree using sequences from exon 2, with top Bayes factors.

and OWM (green) alleles grouping together. Alleles often group by first-field name instead of by species, indicating that some allelic lineages have been maintained since before the split of humans and OWM—at least 31 million years. An example of this is shown in the inset to the left of this tree, "Example of Human-OWM TSP." Here, human alleles coalesce with OWM alleles before they coalesce with each other. Near the bottom of the tree is a clade consisting of ape and NWM MHC-DQB2 sequences, suggesting that they are orthologous; however, NWM species have expanded their MHC-DQ regions so these genes may not be 1:1 orthologous (see our companion paper, ***Fortier and Pritchard*** (***2024***)). Additionally, *Strepsirrhini* sequences do not group with either the MHC-DQB1 or -DQB2 clade, showing that duplications of these genes must have happened in or after the *Simiiformes* ancestor.

Figure 2C shows the *BEAST2* summary tree for non-classical MHC-DOA. In this tree, alleles group exclusively by species (clades are collapsed for clarity) and the branching order of the species deviates only slightly from the species tree. This shows that not all MHC genes are affected by long-term balancing selection, despite the complicated linkage disequilibrium across classical and non-classical genes in the region (***Smith et al., 2024***). This also suggests that antigen presentation specifically, as opposed to a general role in immune function, is the driving force behind this long-term balancing selection.

While the *BEAST2* summary trees in ***Figure 2*** are suggestive of deep TSP, they do not directly quantify the statistical confidence in the TSP model. Moreover, standard approaches to quantifying uncertainty in trees, such as bootstrap support or posterior probabilities for specific clades, do not relate directly to hypothesis testing for TSP. We therefore implemented an alternative approach using *BEAST2* output, as follows (see the Methods for details).

We performed formal model testing for TSP within quartets of alleles, where two alleles are taken from a species (or taxon) A, and two alleles are taken from a different species (or taxon) B. If the alleles from A group together (and the alleles from B group together) in the unrooted tree, this quartet supports monophyly of A (and of B). But if alleles from A group more closely with B in the unrooted tree, then this comparison supports TSP. Since *BEAST2* samples from the posterior distribution of trees, we counted the number of trees that support TSP versus the number that support monophyly as an estimate of the posterior support for each model. We then summarized the relative support for each model by converting these to Bayes factors (see Methods for more detail). The precise interpretation of Bayes factors depends on one's prior expectation; however, following standard guidelines (***Jeffreys, 1998***) we suggest that Bayes factors > 100 should be considered as strong support in favor of TSP. Bayes factors <1 are evidence against TSP. For each comparison of two taxa, we report the maximum Bayes factor across the possible quartets, as we are interested in whether *any* quartet shows compelling evidence for TSP.

Gene conversion, the unidirectional transfer of short tracts of DNA from a donor to an acceptor sequence, can affect the inferred trees. In particular, acceptor sequences may group more strongly with donor sequences than with sequences that share DNA by descent. This can make it difficult to distinguish trees influenced by trans-species polymorphism from those influenced by gene conversion. Thus, we inferred gene conversion tracts using *GENECONV* (***Sawyer, 1999***) and excluded significant gene-converted acceptor alleles from the Bayes factor calculations. While *GENECONV* cannot possibly infer all past events, this procedure should ameliorate any biasing effects. Additionally, we consider each exon separately; analyzing short tracts reduces the effect of recombination on the tree (see our companion paper for more specifics (***Fortier and Pritchard, 2024***)).

Bayes factors are shown in ***Figure 3*** and ***Figure 4***. For each genic region (x-axis), we tested for TSP among humans, non-human primates, and non-primates with each of the comparisons listed on the y-axis. Each table entry is colored and labeled with the maximum Bayes factor among all tested quartets of alleles belonging to that category. See ***Figure 2—figure Supplement 1-Figure 2—figure Supplement 16***, ***Figure 3—figure Supplement 1-Figure 3—figure Supplement 12***, and ***Figure 4—figure Supplement 1-Figure 4—figure Supplement 10*** for examples of high-Bayes-factor

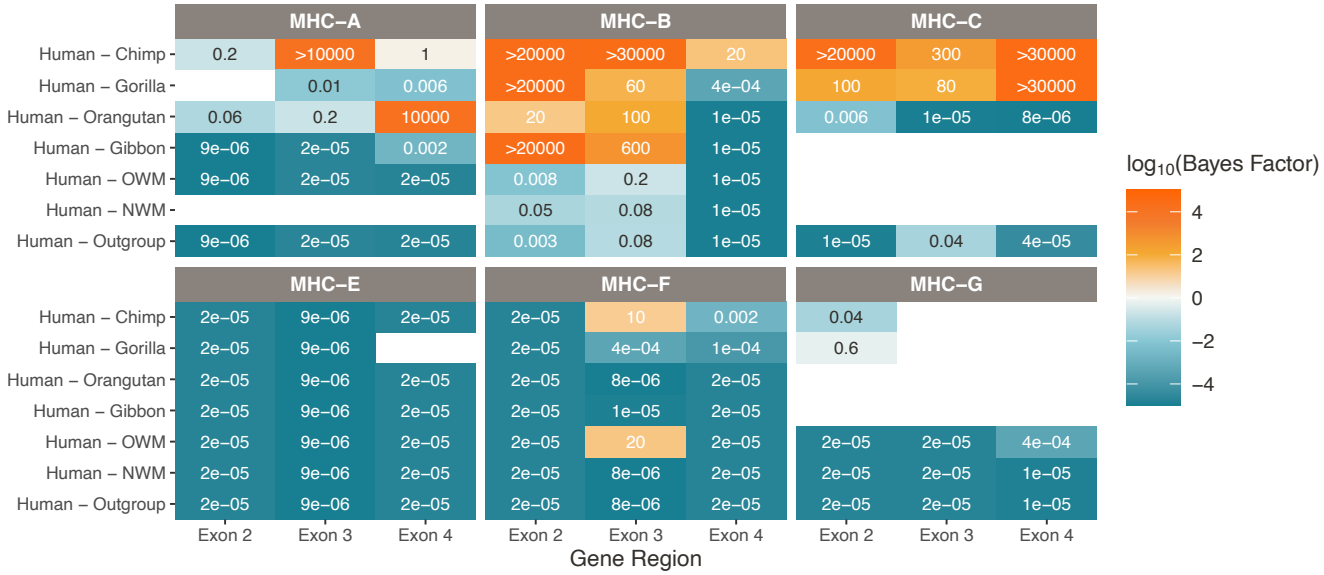

Figure 3. Strong support for TSP at Class I genes MHC-B and -C. Bayes factors computed over the set of BEAST trees indicate deep TSP. Different species comparisons are listed on the y-axis, and different gene regions are listed on the x-axis. High Bayes factors (orange) indicate support for TSP among the given species for that gene region, while low Bayes factors (teal) indicate that alleles assort according to the species tree, as expected. Bayes factors above 100 are considered decisive. Tan values show poor support for either hypothesis, while white boxes indicate that there are not enough alleles in that category with which to calculate Bayes factors. MHC-A is not present in the NWMs, and MHC-C was not present before the human-orangutan ancestor, so it is not possible to calculate Bayes factors for these species comparisons.

Figure 3—figure supplement 1. MHC-A-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 2. MHC-A-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 3. MHC-B-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 4. MHC-B-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 5. MHC-C-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 6. MHC-C-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 7. MHC-E-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 8. MHC-E-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 9. MHC-F-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 10. MHC-F-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 11. MHC-G-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 3—figure supplement 12. MHC-G-Related *BEAST2* tree using sequences from exon 4, with top Bayes factors.

Figure 3—figure supplement 13. Support for TSP between OWM clades for the Class I genes.

Figure 3—figure supplement 14. Support for TSP between NWM clades for the Class I genes.

quartets for each comparison. In ***Figure 3*** and ***Figure 4***, orange indicates support for TSP among the species in that category, while teal indicates evidence against TSP.

At the Class I genes (***Figure 3***), MHC-C shows strong support for TSP within the African apes: human, chimpanzee, and gorilla. Having arisen fairly recently in the ancestor of human and orangutan, MHC-C has thus maintained some allelic lineages for most of its history. TSP has not previously been reported for this gene.

For MHC-A, Bayes factors vary considerably depending on exon and species pair. Many sequences had to be excluded from MHC-A comparisons because they were identified as gene-converted in the *GENECONV* analysis or were previously identified as recombinants (***Hans et al., 2017***; ***Gleimer et al., 2011***; ***Adams and Parham, 2001***). Importantly, for MHC-A we do not see concordance in Bayes factors across the different exons, whereas we do for the other gene groups. Similar Bayes factors across all exons for a given comparison is thus evidence in favor of TSP being the primary driver of the observed deep coalescence structure (rather than recombination or gene conversion).

The non-classical Class I genes MHC-E, -F, and -G (bottom row of ***Figure 3***) are interspersed with the classical Class I genes in the MHC region, but their products have niche functions in innate immunity. Their indirect involvement in adaptive immunity means they experience different selective pressures. They exhibit lower polymorphism and $dN/dS < 1$, reflecting the fact that they have not been subject to the same pathogen-mediated balancing selection. The Bayes factors for all three of these genes show strong evidence against TSP, as expected. Interestingly, despite its non-classical role, MHC-E has a known balanced polymorphism in humans; the two main alleles are at similar frequencies worldwide but may have different expression levels and peptide preferences (***Paganini et al., 2019***; ***Grant et al., 2020***). Our approach—meant to detect ancient TSP—does not reveal balancing selection in MHC-E, showing that this balanced polymorphism is young. For MHC-G, there were not enough sequences available to perform many of the tests (at least two from each species group are required). However, we do not expect to see evidence of TSP in this gene.

Each Class II MHC molecule has an α and β component which are encoded by an A and B gene, respectively. Bayes factors for the Class IIA genes are shown in the top row of ***Figure 4***, while those for their Class IIB partners are shown in the bottom row. The non-classical MHC-DM and -DO molecules assist the classical Class II genes with peptide loading and are not believed to be shaped by balancing selection (***Heijmans et al., 2020***; ***Neefjes et al., 2011***). As expected, we see strong evidence against TSP between humans and all other primate species for these genes (first two columns of ***Figure 4***).

In contrast, we find evidence for deep TSP within the classical Class II genes. MHC-DPA1 shows TSP between human, chimpanzee, and gorilla in exon 2, but not in exon 3. We find that this TSP in MHC-DPA1 is less deep than has been previously suggested with non-Bayesian methods (***Otting and Bontrop, 1995***), underscoring the importance of this methodology for handling the MHC. Meanwhile, its partner MHC-DPB1 shows strong evidence for TSP between human and orangutan in exon 3 and suggestive evidence in exon 2; our work provides the first evidence of TSP between humans and other apes for this gene (***Slierendregt et al., 1995***).

The MHC-DR genes behave somewhat differently than the other classical Class II molecules. While the α and β components of all the other molecules engage in exclusive binding, there are many different MHC-DRβ molecules which all bind to the same MHC-DRα. The MHC-DRA gene is conserved across species with little polymorphism, while the MHC-DRB region is highly variable both in gene content and allelic diversity. Consistent with this, Bayes factors for the MHC-DRA gene reveal strong evidence against TSP for all species pairs, while MHC-DRB1 shows strong evidence in favor of TSP between human, chimpanzee, gorilla, OWM, and even NWM in exon 2. However, because the Bayes factors only support TSP between humans and OWM/NWM in MHC-DRB1 in exon 2, but not in exon 3, this could mean alleles are not actually that ancient. We show in our companion paper that individual MHC-DRB genes are short-lived, and only three are truly orthologous between apes and OWM (***Fortier and Pritchard, 2024***). These pieces of evidence suggest previous work may have overstated the extent of TSP at this locus.

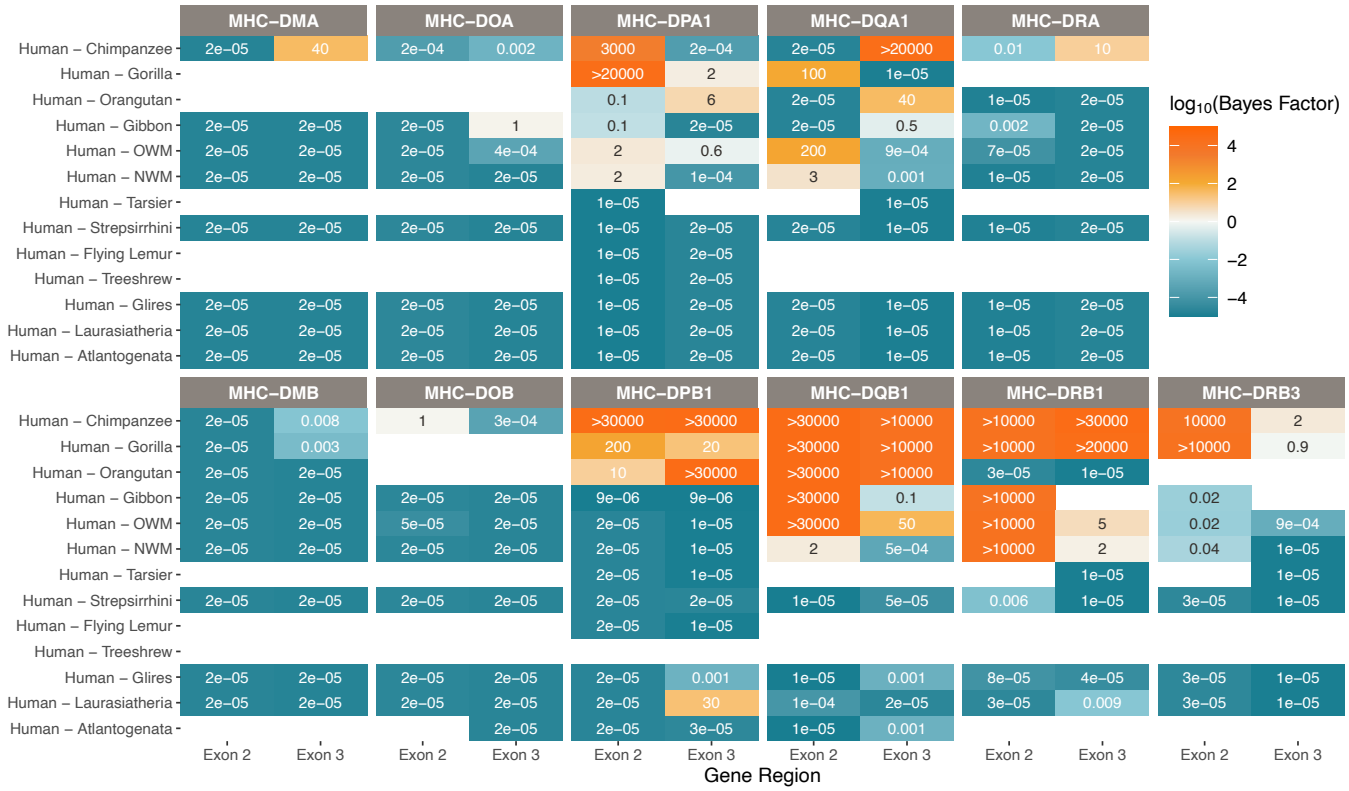

Figure 4. Strong support for TSP at the classical Class II genes. Bayes factors computed over the set of BEAST trees indicate deep TSP. Different species comparisons are listed on the y-axis, and different gene regions are listed on the x-axis. High Bayes factors (orange) indicate support for TSP among the given species for that gene region, while low Bayes factors (teal) indicate that alleles assort according to the species tree, as expected. Bayes factors above 100 are considered decisive. Tan values show poor support for either hypothesis, while white boxes indicate that there are not enough alleles in that category with which to calculate Bayes factors.

Figure 4—figure supplement 1. MHC-DRA-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 2. MHC-DQA-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 3. MHC-DPA-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 4. MHC-DMA-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 5. MHC-DOA-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 6. MHC-DRB-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 7. MHC-DQB-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 8. MHC-DPB-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 9. MHC-DMB-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 10. MHC-DOB-Related *BEAST2* tree using sequences from exon 3, with top Bayes factors.

Figure 4—figure supplement 11. Support for TSP between OWM clades for the Class II genes.

Figure 4—figure supplement 12. Support for TSP between NWM clades for the Class II genes.

Remarkably, the MHC-DQB1 gene shows definitive evidence for TSP back to at least the ancestor of humans and OWM. While this result has been presented previously, we confirm it with decisive evidence (Bayes factor >100) for allelic lineages being maintained for over 31 million years (***Otting et al., 1992***, ***2002***; ***Loisel et al., 2006***; ***Simons et al., 2017***).

Many of our *BEAST* trees also showed intermingling of sequences from different species within the OWM or NWM (e.g. ***Figure 2—figure Supplement 4***), regardless of whether they formed trans-species clades with ape sequences. This could indicate trans-species polymorphism *within* the OWM or NWM that is still ancient, but not old enough to be shared with the apes. Therefore, we also calculated Bayes factors across different clades of OWM and NWM (***Figure 3—figure Supplement 13***, ***Figure 3—figure Supplement 14***, ***Figure 4—figure Supplement 11***, and ***Figure 4—figure Supplement 12***). We see very strong evidence for TSP between all groups of OWM for MHC-DPA1, -DPB1, -DQA1, and -DQB1, indicating that allelic lineages at these genes have been maintained within the OWM for at least 19 million years (MY). Unexpectedly, we also see evidence for TSP in some non-classical genes. MHC-E, a gene which is non-classical in humans and is presumed non-classical in OWM (yet is duplicated in some species), shows evidence for 15-MY-old TSP within the OWM. In non-classical MHC-DMB, we also observe TSP within the OWM as old as 11 MY. Due to the uncertainty of locus assignments for alleles of the OWM MHC-A, -B, and -DRB genes and of the NWM genes, we cannot make definitive conclusions about TSP within these groups for these other genes.

In summary, the phylogenetic analyses point to ancient TSP in classical genes MHC-B, -C, -DPA1, -DPB1, -DQB1, -DRB1, and -DRB3. Bayes factors for the non-classical genes MHC-E, -F, -G, -DMA, -DMB, -DOA, and -DOB suggest there is no TSP involving apes at these loci—as expected for non-classical genes—confirming the usefulness of our approach. However, there may still be TSP at these genes within other groups, such as the OWM, hinting at possible functional differences. Overall, TSP is more ancient among the Class II genes than the Class I genes, consistent with their older age.

From Evolution to Function.

Alongside the evidence for ancient TSP, the MHC region is also notable for its high rate of missense substitutions ($dN/dS > 1$) (***Hughes and Nei, 1988***, ***1989***) and its large number of GWAS hits for autoimmune and infectious diseases (***Buniello et al., 2019***; ***Kennedy et al., 2017***). We next aimed to understand how these observations relate to signals of TSP and known features of the MHC proteins.

To explore these questions, we first estimated the per-site evolutionary rates within each gene. As in our TSP analysis, we used the *BEAST2* package *SubstBMA*, which estimates evolutionary rates at every site concurrently with a tree. We averaged these rates over all states in the chain to get per-site evolutionary rates for each site, then calculated their fold-change relative to the average rate among mostly-gap sites in the alignment ("baseline"; see Methods).

Figure 5A shows the substitution rate fold-change for each nucleotide along the concatenated coding sequence of Class I genes MHC-B, -C, and -E. Exon boundaries are indicated by orange vertical dashed lines. In classical genes MHC-B and -C, nearly all the rapidly-evolving sites lie within exons 2 and 3, which encode the protein's peptide-binding domain. While exons 2 and 3 make up only ∼50% of the gene's length, they contain 94% and 90% of the sites evolving at more than 4 times the baseline evolutionary rate for the classical genes MHC-B and MHC-C, respectively. In MHC-B, exons 2 and 3 each show significantly higher proportions of rapidly-evolving sites compared to the "other" exons (exons in the gene excluding 2, 3, or 4), while the difference is not significant for MHC-C (***Figure 5—figure Supplement 2***). This result could reflect the relatively young age of MHC-C or its additional role as the dominant Class I molecule for interacting with KIRs (***Adams and Parham, 2001***; ***Guethlein et al., 2015***; ***Vollmers et al., 2021***; ***Piontkivska and Nei, 2003***).

In contrast to these classical genes, non-classical MHC-E primarily presents self-peptides for recognition by NK cell receptors, and its peptide-binding groove is tailored to accommodate a very

specific set of self-peptides—leader peptides cleaved from other Class I MHC proteins during processing (***Miller et al., 2003***). As shown in ***Figure 5***A, this gene shows fewer rapidly-evolving sites than the classical genes. These sites are also relatively evenly distributed across the gene, with exons 2 and 3 (which cover ∼50% of the gene's length) containing 45% of the sites evolving at over 4 times the baseline evolutionary rate. Interestingly, exon 4—a non-peptide-binding exon of equal size—displays a significantly lower proportion of rapidly-evolving sites compared with the "other" exons (***Figure 5—figure Supplement 2***). These results support that MHC-E has been remarkably conserved across the primates and that its evolution may not be driven by differential peptide binding (***Heijmans et al., 2020***).

We then examined where the rapidly-evolving sites lie within the physical protein structures. To do this, we averaged the per-site rates within each codon to get per-amino-acid rates, then mapped these onto the known human protein structures. Unfortunately, there are few non-human primate protein structures in the Protein Data Bank (***Berman et al., 2000***), but the macaque structures we found were nearly identical to those of human. ***Figure 5***B shows structures for human HLA-B, -C, and -E; this view features the peptide (black) sitting in the peptide-binding groove (flanked on top and bottom by helices). In MHC-B and -C, rapidly-evolving amino acids (orange) tend to be located within the peptide-binding groove. To quantify this, we measured the minimum distance between each amino acid and the bound peptide. We averaged these distances over several structures, which are listed in the Methods. For all three proteins, amino acids closer to the peptide have significantly higher evolutionary rates than amino acids further from the peptide, as shown in ***Figure 5***C (see also ***Figure 5—figure Supplement 4*** and ***Figure 5—figure Supplement 5***). The effect is much less pronounced in MHC-E, where even the amino acids closest to the peptide do not exhibit high evolutionary rates. These results are consistent with the expectation that rapid evolution and diversity at the classical MHC genes would be mediated by selective pressures for changes in peptide binding.

The rapidly-evolving sites for the Class II genes are shown in ***Figure 6***. Panel A shows the substitution rate fold-change for each nucleotide along the concatenated coding sequence of the Class II MHC-DRA, -DQA, -DRB, and -DQB gene groups. Both MHC-DQA and -DQB are extraordinarily polymorphic, but MHC-DRA is conserved compared to its multiple, highly-variable MHC-DRB partners. Indeed, ***Figure 6***A shows that rapidly-evolving sites are concentrated in binding-site-encoding exon 2 for MHC-DRB, -DQB, and to a lesser extent MHC-DQA. Exon 2, which makes up ∼30% of the coding region, contains 32% of sites evolving at more than twice the baseline rate in MHC-DRA, but 57% of such sites in MHC-DQA, 61% in MHC-DQB, and 73% in MHC-DRB. Comparing across exons, exon 2 contains a significantly higher proportion of rapidly-evolving sites compared to the "other" exons in classical MHC-DQA, -DQB, -DRB, and -DPB, but also—curiously—in non-classical MHC-DMA and -DMB. It is interesting that MHC-DM appears to be evolving rapidly in its binding-site-encoding exons, despite the fact that it is not thought to bind peptides. Instead, it is responsible for assisting with peptide loading onto the classical genes. Co-evolution with MHC-DR in particular seems possible; the interaction between the MHC-DM and -DR molecules depends on the affinity of the peptide trying to bind with MHC-DR. MHC-DM thus shapes the repertoire of peptides presented by MHC-DR, favoring high-affinity peptides (***Dijkstra and Yamaguchi, 2019***; ***Schulze and Wucherpfennig, 2012***). It is plausible that the host-pathogen evolution shaping the MHC-DRB genes has resulted in co-evolution of MHC-DMA and -DMB to maintain this regulatory interaction.

We again mapped the evolutionary rates onto human protein structures, shown in ***Figure 6***B. In each molecule, the α chain is positioned at the top and encompasses the upper helix forming the binding site, while the β chain is oriented toward the bottom and encompasses the lower helix. The peptide is shown in black. Rapidly-evolving sites are concentrated in each protein's binding site, although in MHC-DR this is more prominent in the bottom helix (MHC-DRB).

We then measured the distance between each amino acid and the bound peptide, shown in ***Figure 6***C. MHC-DRA did not show a significant relationship between evolutionary rate and distance, as expected by its relatively uniform distribution of evolutionary rates. For the other three pro-

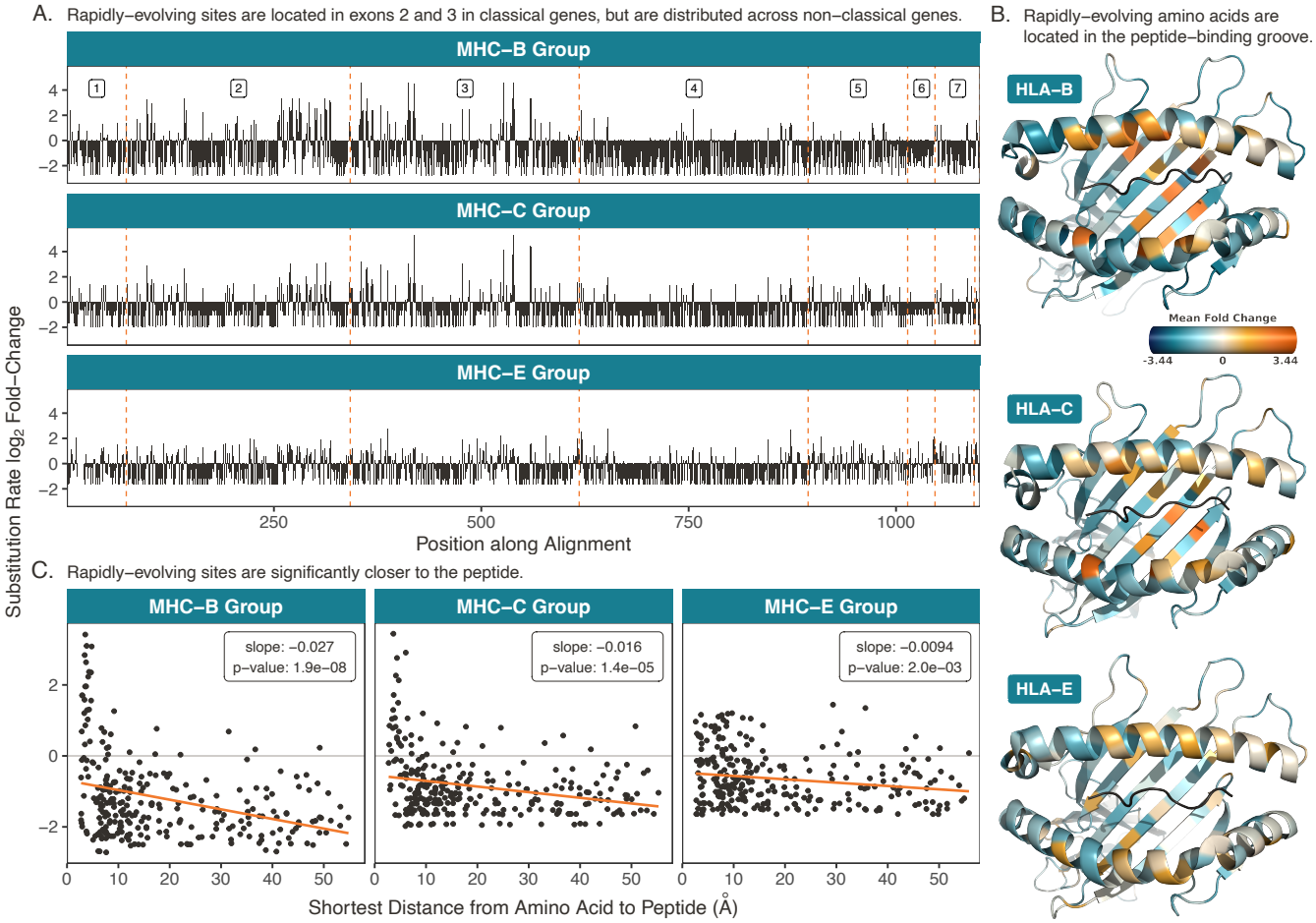

Figure 5. Rapidly-evolving sites in the Class I genes. A) Rapidly-evolving sites are primarily located in exons 2 and 3. Here, the exons are concatenated such that the cumulative position along the coding region is on the x-axis. The dashed orange lines denote exon boundaries. The three genes are aligned such that the same vertical position indicates an evolutionarily equivalent site. The y-axis shows the substitution rate at each site, expressed as a fold-change (the base-2 logarithm of each site's evolutionary rate divided by the mean rate among mostly-gap sites in each alignment; see Methods). **B)** Rapidly-evolving sites are located in each protein's peptide-binding pocket. Structures are Protein Data Bank (***Berman et al., 2000***) 4BCE (***Teze et al., 2014***) for HLA-B, 4NT6 (***Choo et al., 2014***) for HLA-C, and 7P4B (***Walters et al., 2022***) for HLA-E, with images created in *PyMOL* (***Sch, 2021***). Substitution rates for each amino acid are computed as the mean substitution rate of the three sites composing the codon. Orange indicates rapidly-evolving amino acids, while teal indicates conserved amino acids. **C)** Rapidly-evolving amino acids are significantly closer to the peptide than conserved amino acids. The y-axis shows the *BEAST2* substitution rate and the x axis shows the minimum distance to the bound peptide, measured in *PyMOL* (***Sch, 2021***). Each point is an amino acid, and distances are averaged over several structures (see Methods). The orange line is a linear regression of substitution rate on minimum distance, with slope and p-value annotated on each panel.

Figure 5—figure supplement 1. Rapidly-evolving sites along the coding region for all Class I genes.

Figure 5—figure supplement 2. Proportions of rapidly-evolving sites within each exon for Class I.

Figure 5—figure supplement 3. Rapidly-evolving sites mapped to Class I protein structures.

Figure 5—figure supplement 4. Evolutionary rate as a function of minimum distance to peptide for all Class I and Class II proteins.

Figure 5—figure supplement 5. Class I rapidly-evolving sites by binned distance to peptide.

teins, amino acids closer to the peptide had significantly higher evolutionary rates. This held true for the classical MHC-DPA and -DPB genes as well (***Figure 5—figure Supplement 4***). Again, this is consistent with differential peptide binding and TCR responsiveness driving the diversity and long-term balancing selection at the classical genes. We could not measure distances for non-classical MHC-DOA, -DOB, -DMA, and -DMB genes because they do not engage in peptide presentation.

Lastly, since the rapidly-evolving sites are likely involved in peptide binding, they also influence the response to pathogens and self-antigens, presumably affecting risk for infectious and autoimmune diseases. To bridge the gap between evolution and complex traits, we collected HLA fine-mapping studies for infectious, autoimmune, and other diseases, as well as for biomarkers and TCR phenotypes. These studies report associations between a disease or trait and classical HLA alleles, SNPs, and amino acid variants, often with multiple independent hits per gene. They demonstrate that HLA variation affects disease in complex ways—sometimes, a single variant is strongly associated with a condition, while other times, a combination of amino acids or even entire alleles are the strongest indicators of disease susceptibility or protection.

We were interested in whether our rapidly-evolving amino acids from the *BEAST2* analysis corresponded with disease-associated amino acids from the literature. ***Table 1*** lists disease, trait, and TCR-phenotype associations for the most rapidly-evolving amino acids (fold-change ≥ 1) of the MHC-B group (see ***Table 1—source data 1*** for the other genes). The majority of rapidly-evolving positions in MHC-B have at least one association.

Thus, in summary, we find that rapid evolution has primarily targeted amino acids within the peptide-binding region of each gene, and that these specific positions are likely the primary drivers of phenotypic associations at the MHC locus.

Discussion

The MHC region contains the clearest signals of balancing and directional selection in mammalian genomes, including extreme diversity, ancient trans-species polymorphism, and high rates of non-synonymous evolution between allelic lineages. In humans, MHC/HLA variation is associated with risk for infectious and autoimmune diseases and many other traits, and HLA matching is critical for successful tissue transplantation (***Kennedy et al., 2017***; ***Smith et al., 2024***; ***Lee et al., 2007***).

Despite its evolutionary and clinical importance, the extreme diversity of the MHC makes it challenging to study, and basic questions about its evolutionary history remain unresolved. Although past work has suggested ultra-deep TSP at this locus, in this study we re-examined the region with modern, comprehensive data and a unified analysis framework. Using Bayesian evolutionary analysis, we report conclusive evidence for long-term TSP in seven classical genes, including between humans and OWM at MHC-DQB1. Thus, remarkably, allelic lineages at this gene have been maintained for at least 31 million years.

Our evidence for TSP at MHC-DQB1 spanning at least 31 million years places it among the most ancient examples of balancing selection known in any species, and almost certainly the oldest in primates. Aside from MHC, the deepest example within primates is at the ABO locus controlling blood type; it exhibits trans-species polymorphism between humans and gibbons, an age of 23 million years (***Ségurel et al., 2012***). In various chimpanzee species, *OAS1*, which helps inhibit viral replication, contains alleles up to 13MY old (***Ferguson et al., 2012***). TSP between chimpanzee and human includes *LAD1*, a protein that maintains cell cohesion (6MY) (***Teixeira et al., 2015***), retroviral transcription factor *TRIM5α* (4-7MY in apes and > 8MY in OWM) (***Cagliani et al., 2010***; ***Newman et al., 2006***), and *ZC3HAV1*, an antiviral protein leading to viral RNA degradation (6MY) (***Cagliani et al., 2012***), among others (***Leffler et al., 2013***).

Looking more broadly across the tree of life, ancient trans-species polymorphism occurs widely, albeit rarely. Several of the oldest examples are found in the MHC locus: MHC polymorphisms have been maintained for 35MY in cetaceans (***Xu et al., 2009***), 40MY in herons (***Li et al., 2011***), 48MY in mole rats (***Kundu and Faulkes, 2007***), 70MY in tree frogs (***Zhao et al., 2013***), and over

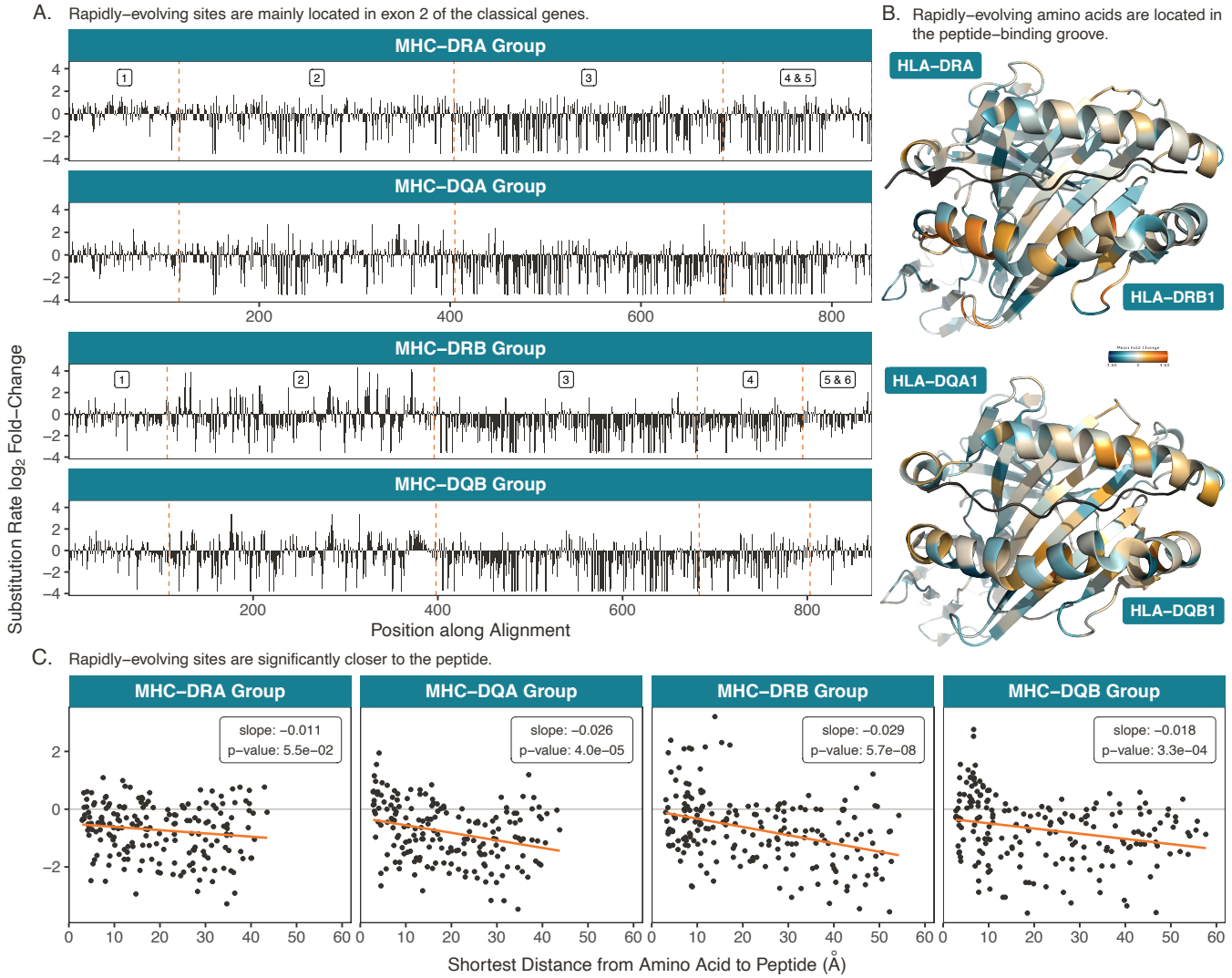

Figure 6. Rapidly-evolving sites in the Class II genes. A) Rapidly-evolving sites are primarily located in exon 2. Here, the exons are concatenated such that the cumulative position along the coding region is on the x-axis. The dashed orange lines denote exon boundaries. The α genes (top two plots) are aligned such that the same vertical position indicates an evolutionarily equivalent site; the same is true for the β genes (bottom two plots). The y-axis shows the substitution rate at each site, expressed as a fold-change (the base-2 logarithm of each site's evolutionary rate divided by the mean rate among mostly-gap sites in each alignment; see Methods). **B)** Rapidly-evolving sites are located in each protein's peptide-binding pocket. Structures are Protein Data Bank (***Berman et al., 2000***) 5JLZ (***Gerstner et al., 2016***) for HLA-DR and 2NNA (***Henderson et al., 2007***) for HLA-DQ, with images created in *PyMOL* (***Sch, 2021***). Substitution rates for each amino acid are computed as the mean substitution rate of the three sites composing the codon. Orange indicates rapidly-evolving amino acids, while teal indicates conserved amino acids. **C)** Rapidly-evolving amino acids are significantly closer to the peptide than conserved amino acids. The y-axis shows the *BEAST2* substitution rate and the x axis shows the minimum distance to the bound peptide, measured in *PyMOL* (***Sch, 2021***). Each point is an amino acid, and distances are averaged over several structures (see Methods). The orange line is a linear regression of substitution rate on minimum distance, with slope and p-value annotated on each panel.

Figure 6—figure supplement 1. Rapidly-evolving sites along the coding region for all Class IIA genes.

Figure 6—figure supplement 2. Proportions of rapidly-evolving sites within each exon for Class IIA.

Figure 6—figure supplement 3. Rapidly-evolving sites along the coding region for all Class IIB genes.

Figure 6—figure supplement 4. Proportions of rapidly-evolving sites within each exon for Class IIB.

Figure 6—figure supplement 5. Rapidly-evolving sites mapped to Class II protein structures.

Figure 6—figure supplement 6. Class II rapidly-evolving sites by binned distance to peptide.

Table 1. Rapidly-evolving amino acids in MHC-B and their trait and disease associations. Shown here are all amino acid positions in the MHC-B group evolving at more than twice the baseline rate (fold-change ≥ 1). Many corresponding positions in human HLA-B have associations with autoimmune or infectious diseases, biomarkers, or TCR phenotypes. Disease associations were collected from a literature search of HLA fine-mapping studies with over 1,000 cases (see Methods).

Amino Acid Position	Evol. Rate Fold-Change	Distance to Peptide (Å)	Associations
156	3.42	3.55	Chronic Hepatitis C (***Hirata et al., 2019***), HIV Set Point Viral Load (***Luo et al., 2021***), Asthma (***Sakaue et al., 2021***), Eosinophil Count (***Sakaue et al., 2021***), Hypothyroidism (***Sakaue et al., 2021***), Pediatric Asthma (***Sakaue et al., 2021***), Systolic Blood Pressure (***Sakaue et al., 2021***), Total Protein (***Sakaue et al., 2021***), TCR β Interaction Probability >50% (***Sharon et al., 2016***), Plasma Protein Levels of ADAM8, AGER, ASPSCR1, B2M, CCL16, CCL28, CCL4, CD200R1, CD5L, CDSN, CX3CL1, FCRL5, IGF2R, IL12A, IL12B, IL5RA, MICB, NUCB2, PDCD1, RARRES2, SFTPD, SIGLEC6, SNX2, TIMD4, TNFRSF4, TNR, TYRP1 (***Krishna et al., 2024***)
95	3.10	3.82	KLRF1 Plasma Protein Level (***Krishna et al., 2024***)
114	3.08	4.80	Rheumatoid Arthritis (***Sakaue et al., 2021***), Plasma Protein Levels of AIF1, CD1C, DDR1, IL15, LILRB2, MICB (***Krishna et al., 2024***)
116	2.84	3.44	Eosinophil Count (***Hirata et al., 2019***), HIV Control (***McLaren et al., 2012***), Angina (***Sakaue et al., 2021***), Allergic Rhinitis (***Waage et al., 2018***), Psoriasis (***Zhou et al., 2016***), Plasma Protein Levels of ADAM15, APOM, BTN2A1, CD1C, CFB, CXCL11, CXCL9, FLT4, GNLY, KLRF1, LILRB1, MICB, PLXDC2, TNF, TNFRSF13C, TNXB (***Krishna et al., 2024***)
70	2.64	3.71	Platelet (***Sakaue et al., 2021***), Plasma Protein Levels of CD8A, GZMA, MICB, NRP2 (***Krishna et al., 2024***)
97	2.37	4.29	Ankylosing Spondylitis (***Butler-Laporte et al., 2023***), HIV Set Point Viral Load (***Luo et al., 2021***), HIV Control (***McLaren et al., 2012***), Adult Height (***Sakaue et al., 2021***), Alkaline Phosphatase (***Sakaue et al., 2021***), Body Weight (***Sakaue et al., 2021***), C-reactive Protein (***Sakaue et al., 2021***), IgA Nephritis (***Sakaue et al., 2021***), Mean Arterial Pressure (***Sakaue et al., 2021***), Mean Corpuscular Hemoglobin (***Sakaue et al., 2021***), Pneumonia (***Tian et al., 2017***), Tonsillectomy (***Tian et al., 2017***), Plasma Protein Levels of ADGRE2, BTN3A2, CCL21, CCL3, CD1C, CD8A, CDSN, CPVL, DXO, EBI3, EFCAB14, HBEGF, HCG22, IL12A, IL12B, LILRB2, LRP1, LRPAP1, LTB, LY75, MANF, MANSC1, MICB, OSCAR, PLA2G10, PRTN3, SIGLEC10, STAB2, TEK, TNFSF13, ZNRD2 (***Krishna et al., 2024***)
24	2.35	4.68	Hepatic Cancer (***Sakaue et al., 2021***), Plasma Protein Levels of ADAM15, CXCL10, FCRL6, GZMB, MICB, TNFSF8 (***Krishna et al., 2024***)
163	2.23	4.13	Lung Cancer (Squamous Cell Carcinoma) (***Ferreiro-Iglesias et al., 2018***), Alanine Aminotransferase (***Sakaue et al., 2021***), TCR Expression (TRAV38-1) (***Sharon et al., 2016***), TCR α Interaction Probability >50% (***Sharon et al., 2016***), Plasma Protein Levels of DDR1, GZMA, LILRB1, MICB, NPTX1, SEPTIN3, TEK, WFDC2 (***Krishna et al., 2024***)
67	1.96	3.72	Graves' Disease (***Hirata et al., 2019***), HIV Set Point Viral Load (***Luo et al., 2021***), Asthma (***Sakaue et al., 2021***), Psoriasis (***Stuart et al., 2022***), Plasma Protein Levels of AMBP, C2, CD160, CD28, CD48, CFB , FCRL1, FCRL6, FRZB, GP1BB, LILRB1, LTA, LY96, NID1, SIGLEC9, SORT1, THBD, TNFRSF4, TNFSF13B, TNXB, TP53BP1, VCAM1 (***Krishna et al., 2024***)
152	1.96	3.53	JIA (Oligoarthritis/RF-negative Polyarthritis) (***Hinks et al., 2017***), Plasma Protein Levels of LTBR, MICB, PLXNA4, RARRES2 (***Krishna et al., 2024***)
63	1.71	2.94	HIV Control (***McLaren et al., 2012***), Skin Cancer (***Sakaue et al., 2021***), B2M Plasma Protein Level(***Krishna et al., 2024***)
99	1.70	3.00	Plasma Protein Levels of APOM, CRTAM, DXO, IL15, MICB, OSCAR (***Krishna et al., 2024***)
66	1.58	3.41	TNFSF11 Plasma Protein Level (***Krishna et al., 2024***)
69	1.56	4.69	Parkinson's Disease (***Naito et al., 2021***)
74	1.31	4.30	Chronic Sinusitis (***Sakaue et al., 2021***)
62	1.29	3.89	PGLYRP1 Plasma Protein Level (***Krishna et al., 2024***)
138	1.26	9.18	
9	1.20	3.36	Primary Biliary Cholangitis (***Darlay et al., 2018***), Systemic Lupus Erythematosus (***Molineros et al., 2019***), Rheumatoid Arthritis (***Raychaudhuri et al., 2012***), Hyperthyroidism (***Sakaue et al., 2021***), Monocyte Count (***Sakaue et al., 2021***), Serum Creatinine (***Sakaue et al., 2021***), Psoriasis (***Zhou et al., 2016***), Plasma Protein Levels of CD1C, CX3CL1, IL12B, LIPF, LTA, MICB, PDCD1, RGMA, SGSH, SLAMF7, TNFRSF8 (***Krishna et al., 2024***)
81	1.03	4.04	

Table 1—source data 1. Rapidly-evolving amino acids and their trait and disease associations for all studied genes.

105MY in salmonid fishes (***Kiryu et al., 2005***; ***Grimholt et al., 2015***). There are examples in non-MHC loci as well; in cyanobacteria, polymorphism at the HEP island controlling heterocyst function has been maintained for 74 million years (***Sano et al., 2018***), in plants, S-genes determining self-incompatibility exhibit TSP spanning 36 million years (***Ioerger et al., 1990***; ***Igic et al., 2006***), and in *Formica* ants, alleles at a supergene underlying colony queen number have been maintained for over 30 million years (***Purcell et al., 2021***).

Paradoxically, given the extremely long-lived balancing selection acting in these lineages, many authors have also reported strong directional selection (***Brandt et al., 2018***; ***Nunes et al., 2021***; ***Bhatia et al., 2011***). Indeed, within the phylogeny we find that the most rapidly-evolving codons are substituted at around 2–4-fold the baseline rate. For the classical genes (except MHC-DRA), these rapidly-evolving sites lie within the peptide binding regions of the corresponding proteins, usually very close to the peptide-contact surfaces. This does not hold true for the non-classical genes, supporting the fact that selection at the MHC is mainly driven by the peptide presentation pathway. Non-classical MHC-DMA and -DMB are a surprising exception, showing significantly elevated proportions of rapidly-evolving sites in exon 2 similarly to the classical genes (even though the MHC-DM molecule does not bind peptides). This pattern may be caused by its co-evolution with the classical Class II genes, and more research is needed to address this.

The primary role of classical MHC proteins is to present peptides for T cell recognition; we found that the same rapidly-evolving amino acids are consistently associated with shaping T cell receptor (TCR) repertoires. Moreover, these amino acids are frequently associated with autoimmune and infectious diseases in HLA fine-mapping studies.

Taken together, we begin to see a comprehensive picture of the nature of primate MHC evolution. In response to rapidly-changing pathogen pressures, the PBRs of classical MHC proteins evolve to bind changing pathogen antigens and present them to TCRs. Broad lineages of MHC alleles are maintained over tens of millions of years by strong balancing selection, providing defense against a wide variety of different pathogens. Yet within these lineages, alleles turn over quickly in response to new specific threats. This reconciles evidence for TSP, the presence of thousands of alleles, and the existence of rapidly-evolving sites.

MHC molecules must be trained to detect pathogens with both specificity and sensitivity, and distinguishing self from non-self peptides is a challenging task. As MHC proteins evolve, there is an unavoidable flux between infection defense and autoimmune susceptibility. Additionally, many MHC proteins have roles in both innate and adaptive immunity. As a result, rapidly-evolving amino acids are associated with both infections and autoimmune conditions. In the future, disease studies within other primate species could provide insight into the trajectory of MHC evolution and might reveal evolutionary trade-offs. Perhaps balancing selection has kept the same amino acids disease-relevant across the entire primate evolutionary tree, or maybe the rapid turnover of MHC variation means different primate clades will have different disease associations.

Although the primate MHC has been of interest to evolutionary biologists for more than 30 years, there is still much to be done to more fully document the evolution of the MHC genes within and between species. Moreover, we still have a limited understanding of how sequence changes map to functional differences among alleles, and how these relate to allele-specific profiles of pathogen protection (and autoimmunity risk). However, functional and computational advances will provide key opportunities for progress on these problems (***Radwan et al., 2020***; ***Vizcaíno et al., 2020***).

Methods and Materials

Data

We downloaded MHC allele nucleotide sequences for all human and nonhuman genes from the IPD Database (updated January 2023) (***Barker et al., 2023***; ***Maccari et al., 2017***, ***2020***). To supplement the alleles available in the database, we also collected nucleotide sequences from NCBI using the

Entrez E-utilities with query "histocompatibility AND txidX AND alive[prop]", where X is a taxon of interest.

We wanted to provide "zoomed-in" versions of various subtrees within the multi-gene trees presented in our companion paper (***Fortier and Pritchard, 2024***). Thus, we included more species and more alleles per species than in the original trees. In each tree, we also included a "backbone" of sequences from the overall multi-gene tree to provide context for each expanded clade (lists of alleles provided as Supplementary Files).

For Class I, we expanded the following clades: 1) MHC-A-related genes (MHC-A group), 2) MHC-B-related genes (MHC-B group), 3) MHC-C-related genes (MHC-C group), 4) MHC-E-related genes (MHC-E group), 5) MHC-F-related genes (MHC-F group), and 6) MHC-G-related genes (MHC-G group). For Class IIA, we expanded: 1) MHC-DMA-related genes (MHC-DMA group), 2) MHC-DOA-related genes (MHC-DOA group), 3) MHC-DRA-related genes (MHC-DRA group), 4) MHC-DPA-related genes (MHC-DPA group), and 5) MHC-DQA-related genes (MHC-DQA group). For Class IIB, we expanded: 1) MHC-DMB-related genes (MHC-DMB group), 2) MHC-DOB-related genes (MHC-DOB group), 3) MHC-DRB-related genes (MHC-DRB group), 4) MHC-DPB-related genes (MHC-DPB group), and 5) MHC-DQB-related genes (MHC-DQB group). These sets were inclusive of all orthologs and paralogs of a given human gene across all species we included (see our companion paper for more information (***Fortier and Pritchard, 2024***)). For example, the MHC-A group includes human HLA-A and its 1:1 orthologs in the apes, the expanded MHC-A and -AG paralogs of the OWM, chimpanzee specific Patr-AL, gorilla-specific Gogo/Gobe-OKO, orangutan-specific Poab/Popy-Ap, and pseudogenes MHC-H and -Y. See Supplementary Files for all alleles included within each group.

We aligned each group separately using *MUSCLE* (***Edgar, 2004***) with default settings and manually adjusted, following the alignments we already produced for the multi-gene trees in our companion paper (***Fortier and Pritchard, 2024***).

Nucleotide Diversity

The classical MHC region is defined as chr6:28,510,120-33,480,577 (GRCh38) (***MHC, 2022***). Nucleotide diversity (π) was calculated on the modern human data using *VCFtools (0.1.15)* (***Danecek et al., 2011***). For the entire MHC region (***Figure 1***A), π was calculated in 5000bp sliding windows with a step size of 1000bp. For each gene separately (***Figure 1***B-C and ***Figure 1—figure Supplement 1***), π was calculated in 50bp sliding windows with a step size of 10bp.

Bayesian Phylogenetic Analysis

We constructed phylogenetic trees using *BEAST2* (***Bouckaert et al., 2014***, ***2019***) with package *SubstBMA* (***Wu et al., 2013***). *SubstBMA* implements a spike-and-slab mixture model that simultaneously estimates the phylogenetic tree, the number of site partitions, the assignment of sites to partitions, the nucleotide substitution model, and a rate multiplier for each partition. Since we were chiefly interested in the partitions and their rate multipliers, we used the RDPM model as described by ***Wu et al.*** (***2013***). In the RDPM model, the number of nucleotide substitution model categories is fixed to 1, so that all sites, regardless of rate partition, share the same estimated nucleotide substitution model. This reduces the number of parameters to be estimated and ensures that only evolutionary rates vary across site partitions, reducing overall model complexity. We used an uncorrelated lognormal relaxed molecular clock because in reality, evolutionary rates vary among branches (***Bergeron et al., 2023***).

Priors. For the Dirichlet process priors, we used the informative priors constructed by ***Wu et al.*** (***2013***) for their mammal dataset. This is appropriate because they include several of the same species and their mammals span approximately the same evolutionary time that we consider in our study. We also use their same priors on tree height, base rate distribution, and a Yule process coalescent prior. We did not specify a calibration point—a time-based prior on a node—because we did not expect our sequences to group according to the species tree.

Running *BEAST2*. We aligned sequences across genes and species and ran *BEAST2* on various subsets of the alignment. For the Class I gene groups (MHC-A group, MHC-B group, MHC-C group, MHC-E group, MHC-F group, and MHC-G group), we repeated the analysis for 1) exon 2 only (PBR), 2) exon 3 only (PBR), 3) exon 4 only (non-PBR), and 4) exons 1, 5, 6, 7, and 8 together (non-PBR; "other" exons). For the Class IIA gene groups (MHC-DMA group, MHC-DOA group, MHC-DRA group, MHC-DPA group, and MHC-DQA group), we used 1) exon 2 only (PBR), 2) exon 3 only (non-PBR), and 3) exons 1, 3, 4, and 5 together (non-PBR; "other" exons). For Class IIB gene groups (MHC-DMB group, MHC-DOB group, MHC-DRB group, MHC-DPB group, and MHC-DQB group), we analyzed 1) exon 2 only (PBR), 2) exon 3 only (non-PBR), and 3) exons 1, 3, 4, and 5 together (non-PBR; "other" exons). In the following, each "analysis" is a collection of *BEAST2* runs using one of these sets of exons of a particular gene group.

The xml files we used to run *BEAST2* were based closely on those used for the mammal dataset with the RDPM model and uncorrelated relaxed clock in ***Wu et al.*** (***2013***)(https://github.com/jessiewu/substBMA/blob/master/examples/mammal/mammal_rdpm_uc.xml). Running a model with per-site evolutionary rate categories and a relaxed clock means there are many parameters to estimate. Along with the large number of parameters, the highly-polymorphic and often highly-diverged sequences in our alignments makes it difficult for *BEAST2* to explore the state space. Thus, we undertook considerable effort to ensure good mixing and convergence of the chains. First, we employed coupled MCMC for all analyses. Coupled MCMC is essentially the same as the regular MCMC used in *BEAST2*, except that it uses additional "heated" chains with increased acceptance probabilities that can traverse unfavorable intermediate states and allow the main chain to move away from an inferior local optimum (***Müller and Bouckaert, 2020***). Using coupled MCMC both speeds up *BEAST2* runs and improves mixing and convergence. We used four heated chains for each run with a delta temperature of 0.025. Second, we ran each *BEAST2* run for 40,000,000 states, discarding the first 4,000,000 states as burn-in and sampling every 10,000 states. Third, we ran at least 8 independent replicates of each analysis. The replicates use the exact same alignment, but explore state space independently and thus are useful for improving the effective sample size of tricky parameters. As recommended by *BEAST2*, we examined all replicates in *Tracer* version 1.7.2 (***Rambaut et al., 2018***) to ensure that they were sampling from the same parameter distributions and had reached convergence. We excluded replicates for which this was not true, as these chains were probably stuck in suboptimal state space. Additionally, Tracer provides estimates of the effective sample size (ESS) for the combined set of states from all chosen replicates, and we required that the ESS be larger than 100 for all parameters. If there were fewer than 4 acceptable replicates or if the ESS was below 100 for any parameter, we re-ran more independent replicates of the analysis until these requirements were satisfied. We obtained between 5 and 18 acceptable replicates per analysis (median 8).

For some analyses, computational limitations prevented *BEAST2* from being able to reach 40,000,000 states. In these situations, more replicates (of fewer states) were usually required to achieve good mixing and convergence. The first 4,000,000 states from each run were still discarded as burn-in even though this represented more than 10% of states in these cases.

This stringent procedure ensured that all of the replicates were exploring the same parameter space and were converging upon the same global optimum, allowing the ≥ 4 independent runs to be justifiably combined. We combined the acceptable replicates using *LogCombiner* version 2.6.7 (***Drummond and Rambaut, 2007***), which aggregates the results across all states. We then used the combined results to perform downstream analyses.

Phylogenetic Trees

After combining acceptable replicates, we obtained 12,382 - 64,818 phylogenies per group/gene region (mean 34,499). We used *TreeAnnotator* version 2.6.3 (***Drummond and Rambaut, 2007***) to summarize each set of possible trees as a maximum clade credibility tree, which is the tree that maximizes the sum of posterior clade probabilities. Since *BEAST2* samples trees from the posterior,

one could in principle perform model testing directly from the posterior samples; the complete set of trees can typically be reduced to a smaller 95% credible set of trees representing the "true" tree (***BEA, 2024***). However, given the high complexity of the model space, all our posterior trees were unique, meaning this was not possible in practice. (Since the prior over tree topologies is unstructured, this effectively puts minuscule prior weight on trees with monophyly. Thus, sampling directly from the posterior provides an unacceptably high-variance estimator.)

Gene Conversion

We calculated gene conversion fragments using *GENECONV* version 1.81a (***Sawyer, 1999***) on each alignment. It is generally advisable to use only synonymous sites when running the program on a protein-coding alignment, since silent sites within the same codon position are likely to be correlated. However, the extreme polymorphism in these MHC genes meant there were too few silent sites to use in the analysis. Thus, we considered all sites but caution that this could slightly overestimate the lengths of our inferred conversion tracts. For each alignment, we ran *GENECONV* with options *ListPairs*, *Allouter*, *Numsims* $= 10000$, and *Startseed* $= 310$. We collected all inferred "Global Inner" (GI) fragments with $sim_pval < 0.05$ (this is pre-corrected for multiple comparisons by the program). GI fragments represent a possible gene conversion event between two sequences in the alignment.

For each GI fragment, we made an educated guess on which sequence was the donor sequence and which was the acceptor sequence by comparing how sequences clustered using sites within the fragment bounds to how sequences clustered using sites outside of the fragment bounds (but in the same exon). Sequences that were confidently determined to be acceptor sequences were excluded from the Bayes factor analyses for the relevant exon because their non-tree-like behavior has the potential to bias results. For sequence pairs where the direction could not be determined, both sequences were excluded from subsequent analyses.

Bayes Factors

Because we could not perform model testing directly on the full phylogenies, we used an alternative approach—computing Bayes factors for TSP within manageable subsets of the data, i.e., quartets of alleles. Let D be a sample of phylogenies from *BEAST2*, sampled from the posterior with uniform prior. For a chosen species, we have a null hypothesis H, that human alleles form a monophyletic group, and an alternative hypothesis, H^c, that is also the complement of H—that the human alleles do not form a monophyletic group. The Bayes factor, K, is a ratio quantifying support for the alternative hypothesis:

$$K = \frac{\Pr(D|H^c)}{\Pr(D|H)} = \frac{\Pr(H^c|D)}{\Pr(H|D)} \cdot \frac{\Pr(H)}{\Pr(H^c)}$$

where the first term on the right hand side is the posterior odds in favor of the alternative hypothesis and the second term is the prior odds in favor of the null hypothesis. Bayes factors above 100 are considered decisive support for the alternative hypothesis (***Jeffreys, 1998***).

Because it is difficult to evaluate monophyly using a large number of alleles, we evaluate Bayes factors considering 4 alleles at a time: 2 alleles of a single species and 2 alleles of different species. For example, to assess support for TSP between humans and chimpanzees, we could use 2 human alleles and 2 bonobo alleles. Or, to assess support for TSP between humans and OWM, we could use 2 human alleles, one baboon, and one macaque allele. Because there are many possible sets of 4 alleles for each comparison, we tested a large number of quartets. We reported the *maximum* Bayes factor among all tested allele sets to represent evidence for TSP for that species comparison, because our aim was to find *any* evidence of TSP among *any* set of 4 alleles.

Next, we calculated the prior odds of the null hypothesis (that the chosen species, i.e. humans, form a monophyletic group). The prior odds $\frac{\Pr(H)}{\Pr(H^c)} = \frac{1}{2}$, because if the trees were assembled at random, there is 1 possible unrooted tree where the 2 human alleles would form a monophyletic

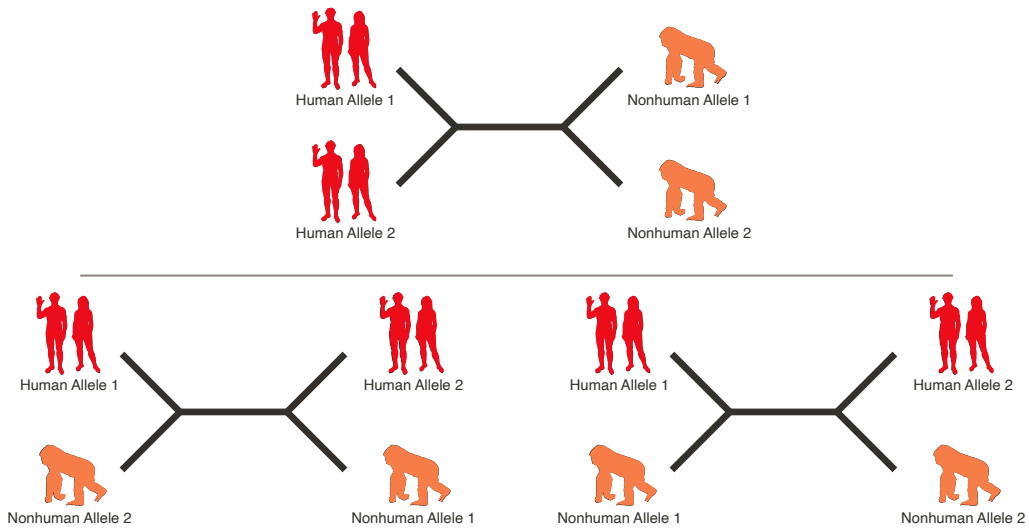

Figure 7. Possible unrooted trees of 4 alleles. There is one tree where the human alleles are monophyletic, and two trees where they are non-monophyletic.

group and 2 possible unrooted trees where the 2 human alleles would not form a monophyletic group, as shown in ***Figure 7***.

The data, D, is the set of *BEAST2* trees, so the posterior odds $\frac{\Pr(H^c|D)}{\Pr(H|D)}$ is the fraction of *BEAST2* trees where the 2 human alleles **do not** form a monophyletic group divided by the fraction of *BEAST2* trees where the 2 human alleles **do** form a monophyletic group. If either fraction is 0, we set its probability to $p = \frac{1}{n+1}$, where n is the number of *BEAST2* trees for that gene/sequence subset, and set the complement's probability to $1 - p$. This is the reason that some labels in ***Figure 3-Figure 4*** contain a > sign (e.g. if no trees in a set of $14,000$ were monophyletic, then the Bayes factor must be at minimum $7,000$).

Bayes factors K were then computed as follows and interpreted according to the scale given by ***Jeffreys*** (***1998***).

$$K = \frac{\Pr(H^c|D)}{\Pr(H|D)} \cdot \frac{1}{2}$$

For each gene and genic region, we tested for TSP between human and chimpanzee, gorilla, orangutan, gibbon, OWM, and NWM. For the Class II genes, which have orthologs beyond the primates, we also tested for TSP between human and tarsier, *Strepsirrhini*, Flying Lemur, Treeshrew, *Glires*, *Laurasiatheria*, and *Atlantogenata*. For the Class I genes, we considered outgroup species to be the tarsier and *Strepsirrhini*.

Rapidly-Evolving Sites

BEAST2 places sites into partitions and estimates evolutionary rates for each partition. We averaged these rates over all sampled states, resulting in an overall evolutionary rate for each nucleotide position. In designing the gene groups, we included common "backbone sequences" in every set, although we expanded one particular clade for each focused tree. The inclusion of genes spanning the whole MHC family caused the alignments to contain many gaps. To normalize the rates, we took advantage of the fact that every alignment had many mostly-gap sites. We defined "gappy" sites as those in which the alignment had a gap in all of the following human backbone alleles: HLA-A*01:01:01:01, HLA-B*07:02:01:01, HLA-C*01:02:01:01, HLA-E*01:01:01:01, HLA-F*01:01:01:01, HLA-G*01:01:01:01, and HLA-J*01:01:01:01 for Class I, HLA-DRA*01:01:01:01, HLA-DQA1*01:01:01:01, HLA-DPA1*01:03:01:01, HLA-DMA*01:01:01:01, and HLA-DOA*01:01:01:01 for Class IIA, and HLA-DRB1*01:01:01:01, HLA-DQB1*02:01:01:01, HLA-DPB1*01:01:01:01, HLA-DMB*01:01:01:0 and HLA-DOB*01:01:01:01 for Class IIB. Because mostly-gap sites are not expected to affect the

BEAST2 run very much and were common to all group alignments, we considered these sites' *BEAST2* evolutionary rates as a baseline. For each group and gene region, we calculated the mean rate among all of these baseline mostly-gap sites. Then, we expressed normalized per-site rates as fold-changes by taking the base-2 logarithm of the ratio of each site's evolutionary rate to the baseline mean.

Protein Structures.

To map the rapidly-evolving sites onto protein structures, we first translated our nucleotide alignments into protein sequences using Expasy translate (***Gasteiger et al., 2003***). We then aligned our translated sequences with amino acid sequences from selected Protein Data Bank (PDB) (***Berman et al., 2000***)(https://www.rcsb.org/) structures (listed below) using *MUSCLE* (***Edgar, 2004***) with default settings.

The translated alignments and the PDB structures' sequences allowed us to determine which nucleotides in the alignment corresponded to which amino acids in each structure. We then calculated per-amino-acid evolutionary rates by averaging the per-site evolutionary rates among the sites composing each codon. We caution that the vast majority of PDB structures we considered were from human (and were functional), and thus they cannot capture all of the indels and structural changes that define different primate proteins. Additionally, some of our alignments included pseudogenes and diverse sets of related genes, for which it may seem reductionist to map to a single human protein structure. However, we note that despite significant sequence variation, the structures of different MHC proteins (from different genes within the same class) are extremely similar, and thus we do not expect such complications to alter the results.

We used *PyMOL* version 2.4.2 (***Sch, 2021***) to visualize the per-amino acid evolutionary rates on each gene's protein structure. We used model 4BCE (***Teze et al., 2014***) for HLA-B, 4NT6 (***Choo et al., 2014***) for HLA-C, 7P4B (***Walters et al., 2022***) for HLA-E, 5JLZ (***Gerstner et al., 2016***) for HLA-DR and 2NNA (***Henderson et al., 2007***) for HLA-DQ from Protein Data Bank (***Berman et al., 2000***) to prepare the main figures (https://www.rcsb.org/).

We calculated the distances between all atoms of all amino acids of the HLA molecule and all atoms of all amino acids of the peptide in *PyMOL*, then took the minimum distance to represent each amino acid's distance to the peptide. Where possible, we averaged the minimum distances over multiple alternative structures for each protein, to avoid relying too heavily on a particular structure. The structures chosen represented different alleles, different contexts (e.g. bound to a receptor or not, bound to a self or non-self peptide), and in a few cases different species.

We used models 1ZVS (***Chu et al., 2007***), 3JTT (***Dai et al., 2010a***), 3OX8 (***Liu et al., 2011***), 3OXR (***Liu et al., 2011***), 3OXS (***Liu et al., 2011***), 3RL2 (***Zhang et al., 2011***), 4HX1 (***Niu et al., 2013***), 6J1V (***Zhu et al., 2019***), 6J1W (***Zhu et al., 2019***), 6MPP (***Flores-Solis et al., 2019***), 6PBH (***van de Sandt et al., 2019***), 7SR0 (***Finton et al., 2023***), 7SRK (***Finton et al., 2023***), 7WT5 (***Asa et al., 2022***), and 8I5C (***Lu et al., 2023***) from Protein Data Bank (***Berman et al., 2000***) for MHC-A; 1JGD (***Hillig et al., 2004***), 3BVN (***Kumar et al., 2009***), 3KPL (***Macdonald et al., 2009***), 3KPN (***Macdonald et al., 2009***), 3LN4 (***Bade-Döding et al., 2011***), 3LN5 (***Bade-Döding et al., 2011***), 3RWJ (***Wu et al., 2011***), 3W39 (***Yagita et al., 2013***), 3X13 (***Saunders et al., 2015***), 4JQV (***Rist et al., 2013***), 4JRY (***Liu et al., 2013***), 4MJI (***Motozono et al., 2014***), 4O2E (***Sun et al., 2014***), 4PRA (***Liu et al., 2014***), 4PRB (***Liu et al., 2014***), 5EO0 (***Du et al., 2016***), 5IEK (***Alpizar et al., 2016***), 5VUD (***Illing et al., 2018***), 5VVP (***Illing et al., 2018***), 5VWF (***Illing et al., 2018***), 6IWG (***Yamamoto et al., 2019***), 6MTM (***Grant et al., 2018***), 6PYL (***Sian et al., 2019***), 6PYV (***Sian et al., 2019***), 6UZP (***Schutte et al., 2020***), 6VIU (***Schutte et al., 2020***), 6Y27 (***Loll et al., 2020***), 7R7V (***Li et al., 2023***), 7T0L (***Vivian and Rossjohn, 2022***), 7TUC (***Jiang et al., 2022b***), 7X1C (***Huan et al., 2023***), and 7YG3 (***Jiang et al., 2022a***) for MHC-B; 4NT6 (***Choo et al., 2014***), 5VGD (***Kaur et al., 2017***), 5VGE (***Kaur et al., 2017***), 5W67 (***Mobbs et al., 2017***), 6PAG (***Moradi et al., 2021***), and 7WJ3 (***Asa et al., 2022***) for HLA-C; 2ESV (***Hoare et al., 2006***), 3CDG (***Petrie et al., 2008***), 5W1V (***Sullivan et al., 2017***), 7P49 (***Walters et al., 2022***), and 7P4B (***Walters et al., 2022***) for HLA-E; 5IUE (***Dulberger et al., 2017***) for HLA-F; 1YDP (***Clements et al., 2005***), 2DYP (***Shiroishi et al., 2006***), and 3KYN (***Walpole et al.,***

2010) for HLA-G; 1HDM (*Mosyak et al., 1998*), 2BC4 (*Nicholson et al., 2006*), 4FQX (*Pos et al., 2012*), 4GBX (*Pos et al., 2012*), and 4I0P (*Guce et al., 2013*) for HLA-DM; 4I0P (*Guce et al., 2013*) for HLA-DO; 3LQZ (*Dai et al., 2010b*), 3WEX (*Kusano et al., 2014*), 7T2A (*Ciacchi et al., 2023*), 7T6I (*Sebastian Klobuch et al., 2022*), and 7ZAK (*Racle et al., 2023*) for HLA-DP; 2NNA (*Henderson et al., 2007*), 4D8P (*Tollefsen et al., 2012*), 5KSA (*Petersen et al., 2016*), 5KSU (*Nguyen et al., 2017*), 6DIG (*Jiang et al., 2019*), and 6PX6 (*Ting et al., 2020*) for HLA-DQ; and 1BX2 (*Smith et al., 1998*), 1FV1 (*Li et al., 2000*), 1H15 (*Lang et al., 2002*), 1T5X (*Zavala-Ruiz et al., 2004*), 2Q6W (*Parry et al., 2007*), 3C5J (*Dai et al., 2008*), 4FQX (*Pos et al., 2012*), 4H1L (*Yin et al., 2012*), 5JLZ (*Gerstner et al., 2016*), 5V4M (*Ooi et al., 2017*), 6ATF (*Scally et al., 2017*), and 8EUQ (*Kassardjian et al., 2023*) for HLA-DR.

Disease Literature.

We conducted a literature search for papers that used HLA fine-mapping to discover disease associations, limiting our selection to those including at least 1,000 cases and which identified putatively independent signals via conditional analysis. We included all independent amino acid signals identified as significant by the original authors. If there was more than one study for the same disease, but in different populations, we included all unique independent hits. We also collected associations between amino acids and TCR phenotypes.

Polymorphic Amino Acid Positions.

Amino acid allele frequencies were obtained from *Luo et al.* (*2021*), who imputed HLA variation in 1000G using *SNP2HLA* (https://github.com/immunogenomics/HLA-TAPAS/blob/master/resources/1000G.bglv4.FRQ.frq).

Acknowledgments

We acknowledge support from NIH grants R01 HG011432 and R01 HG008140. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1656518. We appreciate helpful comments from Jeffrey Spence, the Pritchard lab, and the reviewers of the previous version of this work.

References

The PyMOL Molecular Graphics System, Version 2.4.2. Schrödinger, LLC; 2021. https://pymol.org/2/.

Human Genome Region MHC; 2022. https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC.

SUMMARIZING POSTERIOR TREES. Centre for Computational Evolution; 2024. https://www.beast2.org/summarizing-posterior-trees/.

Abi-Rached L, Kuhl H, Roos C, ten Hallers B, Zhu B, Carbone L, de Jong PJ, Mootnick AR, Knaust F, Reinhardt R, Parham P, Walter L. A Small, Variable, and Irregular Killer Cell Ig-Like Receptor Locus Accompanies the Absence of MHC-C and MHC-G in Gibbons . The Journal of Immunology. 2010; 184(3):1379–1391. doi: 10.4049/jimmunol.0903016.

Adams EJ, Luoma AM. The adaptable major histocompatibility complex (MHC) fold: Structure and function of nonclassical and mhc class I-like molecules. Annual Review of Immunology. 2013; 31:529–561. doi: 10.1146/annurev-immunol-032712-095912.

Adams EJ, Parham P. Species-specific evolution of MHC class I genes in the higher primates. Immunological Reviews. 2001; 183:41–64. doi: 10.1034/j.1600-065X.2001.1830104.x.

Allentoft ME, Sikora M, Refoyo-martínez A, Irving-pease EK, Fischer A. Population Genomics of Stone Age Eurasia. bioRxiv. 2022; doi: 10.1101/2022.05.04.490594.

Alpizar A, Marcilla M, Santiago C. Structure of HLA-B*40:02 in complex with the endogenous peptide REFSKEPEL. PDB. 2016; .

Arden B, Klein J. Biochemical comparison of major histocompatibility complex molecules from different subspecies of Mus musculus: Evidence for trans-specific evolution of alleles. Proceedings of the National Academy of Sciences of the United States of America. 1982; 79(7):2342–2346. doi: 10.1073/pnas.79.7.2342.

Asa M, Morita D, Kuroha J, Mizutani T, Mori N, Mikami B, Sugita M. Crystal structures of N-myristoylated lipopeptide-bound HLA class I complexes indicate reorganization of B-pocket architecture upon ligand binding. Journal of Biological Chemistry. 2022; 298(7):102100. https://doi.org/10.1016/j.jbc.2022.102100, doi: 10.1016/j.jbc.2022.102100.

Azevedo L, Serrano C, Amorim A, Cooper DN. Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. Human Genomics. 2015 12; 9(1):21. http://www.humgenomics.com/content/9/1/21, doi: 10.1186/s40246-015-0043-1.

Bade-Döding C, Theodossis A, Gras S, Kjer-Nielsen L, Eiz-Vesper B, Seltsam A, Huyton T, Rossjohn J, McCluskey J, Blasczyk R. The impact of human leukocyte antigen (HLA) micropolymorphism on ligand specificity within the HLA-B*41 allotypic family. Haematologica. 2011; 96(1):110–118. doi: 10.3324/haematol.2010.030924.

Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, Marsh SGE. The IPD-IMGT/HLA Database. Nucleic Acids Research. 2023; 51(November 2022):1053–1060.

Bergeron LA, Besenbacher S, Zheng J, Li P, Bertelsen MF, Quintard B, Hoffman JI, Li Z, St Leger J, Shao C, Stiller J, Gilbert MTP, Schierup MH, Zhang G. Evolution of the germline mutation rate across vertebrates. Nature. 2023; 615(7951):285–291. doi: 10.1038/s41586-023-05752-y.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Research. 2000; 28:235–242. doi: 10.1093/nar/28.1.235.

Bhatia G, Patterson N, Pasaniuc B, Zaitlen N, Genovese G, Pollack S, Mallick S, Myers S, Tandon A, Spencer C, Palmer CD, Adeyemo AA, Akylbekova EL, Cupples LA, Divers J, Fornage M, Kao WHL, Lange L, Li M, Musani S, et al. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. American Journal of Human Genetics. 2011; 89(3):368–381. doi: 10.1016/j.ajhg.2011.07.025.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Computational Biology. 2014; 10(4):1–6. doi: 10.1371/journal.pcbi.1003537.

Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, Du Plessis L, Popinga A, Rambaut A, Rasmussen D, Siveroni I, Suchard MA, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Computational Biology. 2019; 15(4):1–28. doi: 10.1371/journal.pcbi.1006650.

Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI. The MHC class I genes of the rhesus monkey: Different evolutionary histories of MHC class I and II genes in primates. The Journal of Immunology. 1996; 156(12):4656–4665.

Brändle U, Ono H, Vincek V, Klein D, Golubic M, Grahovac B, Klein J. Trans-species evolution of Mhc-DRB haplotype polymorphism in primates: Organization of DRB genes in the chimpanzee. Immunogenetics. 1992; 36(1):39–48. doi: 10.1007/BF00209291.

Brandt DYC, César J, Goudet J, Meyer D. The effect of balancing selection on population differentiation: A study with HLA genes. G3: Genes, Genomes, Genetics. 2018; 8(8):2805–2815. doi: 10.1534/g3.118.200367.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Research. 2019 1; 47(D1):D1005–D1012. https://academic.oup.com/nar/article/47/D1/D1005/5184712, doi: 10.1093/nar/gky1120.

Butler-Laporte G, Farjoun J, Nakanishi T, Lu T, Abner E, Chen Y, Hultström M, Metspalu A, Milani L, Mägi R, Nelis M, Hudjashov G, Yoshiji S, Ilboudo Y, Liang KYH, Su CY, Willet JDS, Esko T, Zhou S, Forgetta V, et al. HLA allele-calling using multi-ancestry whole-exome sequencing from the UK Biobank identifies 129 novel associations in 11 autoimmune diseases. Communications Biology. 2023; 6(1):1–17. doi: 10.1038/s42003-023-05496-5.

Cagliani R, Fumagalli M, Biasin M, Piacentini L, Riva S, Pozzoli U, Bonaglia MC, Bresolin N, Clerici M, Sironi M. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. Human Genetics. 2010 12; 128(6):577–588. http://link.springer.com/10.1007/s00439-010-0884-6, doi: 10.1007/s00439-010-0884-6.

Cagliani R, Guerini FR, Fumagalli M, Riva S, Agliardi C, Galimberti D, Pozzoli U, Goris A, Dubois B, Fenoglio C, Forni D, Sanna S, Zara I, Pitzalis M, Zoledziewska M, Cucca F, Marini F, Comi GP, Scarpini E, Bresolin N, et al. A Trans-Specific Polymorphism in ZC3HAV1 Is Maintained by Long-Standing Balancing Selection and May Confer Susceptibility to Multiple Sclerosis. Molecular Biology and Evolution. 2012 6; 29(6):1599–1613. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss002, doi: 10.1093/molbev/mss002.

Choo JAL, Liu J, Toh X, Grotenbreg GM, Ren EC. The Immunodominant Influenza A Virus M1 58–66 Cytotoxic T Lymphocyte Epitope Exhibits Degenerate Class I Major Histocompatibility Complex Restriction in Humans. Journal of Virology. 2014; 88(18):10613–10623. doi: 10.1128/jvi.00855-14.

Chu F, Lou Z, Chen YW, Liu Y, Gao B, Zong L, Khan AH, Bell JI, Rao Z, Gao GF. First Glimpse of the Peptide Presentation by Rhesus Macaque MHC Class I: Crystal Structures of Mamu-A*01 Complexed with Two Immunogenic SIV Epitopes and Insights into CTL Escape. The Journal of Immunology. 2007; 178(2):944–952. doi: 10.4049/jimmunol.178.2.944.

Ciacchi L, van de Garde MDB, Ladell K, Farenc C, Poelen MCM, Miners KL, Llerena C, Reid HH, Petersen J, Price DA, Rossjohn J, van Els CACM. CD4+ T cell-mediated recognition of a conserved cholesterol-dependent cytolysin epitope generates broad antibacterial immunity. Immunity. 2023; 56(5):1082–1097. doi: 10.1016/j.immuni.2023.03.020.

Clements CS, Kjer-Nielsen L, Kostenko L, Hoare HL, Dunstone MA, Moses E, Freed K, Brooks AG, Rossjohn J, McCluskey J. Crystal structure of HLA-G: A nonclassical MHC class I molecule expressed at the fetal-maternal interface. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(9):3360–3365. doi: 10.1073/pnas.0409676102.

Cong Pk, Bai WY, Li JC, Yang MY, Khederzadeh S, Gai SR, Li N, Liu YH, Yu SH, Zhao WW, Liu JQ, Sun Y, Zhu XW, Zhao PP, Xia JW, Guan PL, Qian Y, Tao JG, Xu L, Tian G, et al. Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. Nature Communications. 2022 12; 13(1):2939. https://www.nature.com/articles/s41467-022-30526-x, doi: 10.1038/s41467-022-30526-x.

Dai L, Feng Y, Qi J, Gao GF, Crystal structure of Rhesus macaque MHC class I:Mamu-A*02; 2010. https://doi.org/10.2210/pdb3JTT/pdb.

Dai S, Crawford F, Marrack P, Kappler JW. The structure of HLA-DR52c: Comparison to other HLA-DRB3 alleles. Proceedings of the National Academy of Sciences of the United States of America. 2008; 105(33):11893–11897. doi: 10.1073/pnas.0805810105.

Dai S, Murphy GA, Crawford F, Mack DG, Falta MT, Marrack P, Kappler JW, Fontenot AP. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107(16):7425–7430. doi: 10.1073/pnas.1001772107.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCFtools. Bioinformatics. 2011 8; 27(15):2156–2158. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btr330, doi: 10.1093/bioinformatics/btr330.

Darlay R, Ayers KL, Mells GF, Hall LS, Liu JZ, Almarri MA, Alexander GJ, Jones DE, Sandford RN, Anderson CA, Cordell HJ. Amino acid residues in five separate HLA genes can explain most of the known associations between the MHC and primary biliary cholangitis. PLoS Genetics. 2018; 14(12):1–29. doi: 10.1371/journal.pgen.1007833.

De Groot NG, Otting N, Robinson J, Blancher A, Lafont BAP, Marsh SGE, O'Connor DH, Shiina T, Walter L, Watkins DI, Bontrop RE. Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. Immunogenetics. 2012; 64(8):615–631. doi: 10.1007/s00251-012-0617-1.

Dijkstra JM, Yamaguchi T. Ancient features of the MHC class II presentation pathway, and a model for the possible origin of MHC molecules. Immunogenetics. 2019; 71(3):233–249. doi: 10.1007/s00251-018-1090-2.

Douillard V, Castelli EC, Mack SJ, Hollenbach JA, Gourraud PA, Vince N, Limou S. Approaching Genetics Through the MHC Lens: Tools and Methods for HLA Research. Frontiers in Genetics. 2021; 12(December):1–13. doi: 10.3389/fgene.2021.774916.

Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 2007; 7(1):214. http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-7-214, doi: 10.1186/1471-2148-7-214.

Du VY, Bansal A, Carlson J, Salazar-Gonzalez JF, Salazar MG, Ladell K, Gras S, Josephs TM, Heath SL, Price DA, Rossjohn J, Hunter E, Goepfert PA. HIV-1–Specific CD8 T Cells Exhibit Limited Cross-Reactivity during Acute Infection. The Journal of Immunology. 2016; 196(8):3276–3286. doi: 10.4049/jimmunol.1502411.

Dulberger CL, McMurtrey CP, Hölzemer A, Neu KE, Liu V, Steinbach AM, Garcia-Beltran WF, Sulak M, Jabri B, Lynch VJ, Altfeld M, Hildebrand WH, Adams EJ. Human Leukocyte Antigen F Presents Peptides and Regulates Immunity through Interactions with NK Cell Receptors. Immunity. 2017; 46(6):1018–1029. doi: 10.1016/j.immuni.2017.06.002.

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 2004 3; 32(5):1792–1797. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh340, doi: 10.1093/nar/gkh340.

Ferguson W, Dvora S, Fikes RW, Stone AC, Boissinot S. Long-Term Balancing Selection at the Antiviral Gene OAS1 in Central African Chimpanzees. Molecular Biology and Evolution. 2012 4; 29(4):1093–1103. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msr247, doi: 10.1093/molbev/msr247.

Ferreiro-Iglesias A, Lesseur C, McKay J, Hung RJ, Han Y, Zong X, Christiani D, Johansson M, Xiao X, Li Y, Qian DC, Ji X, Liu G, Caporaso N, Scelo G, Zaridze D, Mukeriya A, Kontic M, Ognjanovic S, Lissowska J, et al. Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. Nature Communications. 2018; 9(1):1–12. doi: 10.1038/s41467-018-05890-2.

Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, Yengo L, Rocheleau G, Froguel P, McCarthy MI, Pritchard JK. Detection of human adaptation during the past 2000 years. Science. 2016; 354(6313):760–764. doi: 10.1126/science.aag0776.

Figueroa F, Gúnther E, Klein J. MHC polymorphism pre-dating speciation. Nature. 1988 9; 335(6187):265–267. http://www.nature.com/articles/335265a0, doi: 10.1038/335265a0.

Finton KAK, Rupert PB, Friend DJ, Dinca A, Lovelace ES, Buerger M, Rusnac DV, Foote-McNabb U, Chour W, Heath JR, Campbell JS, Pierce RH, Strong RK. Effects of HLA single chain trimer design on peptide presentation and stability. Frontiers in Immunology. 2023; 14(May):1–11. doi: 10.3389/fimmu.2023.1170462.

Flores-Solis D, McShan AC, Sgourakis NG, HLA-A*01:01 complex with NRAS Q61K peptide by NMR; 2019. https://doi.org/10.2210/pdb6MPP/pdb.

Foley NM, Mason VC, Harris AJ, Bredemeyer KR, Damas J, Lewin HA, Eizirik E, Gatesy J, Karlsson EK, Lindblad-Toh K, Springer MS, Murphy WJ, Andrews G, Armstrong JC, Bianchi M, Birren BW, Bredemeyer KR, Breit AM, Christmas MJ, Clawson H, et al. A genomic timescale for placental mammal evolution. Science. 2023; 380(6643). doi: 10.1126/science.abl8189.

Fortier AL, Pritchard JK. The Primate Major Histocompatibility Complex: An Illustrative Example of Gene Family Evolution. TBD. 2024; .

Fukami-Kobayashi K, Shiina T, Anzai T, Sano K, Yamazaki M, Inoko H, Tateno Y. Genomic evolution of MHC class I region in primates. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(26):9230–9234. doi: 10.1073/pnas.0500770102.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Research. 2003; 31(13):3784–3788. doi: 10.1093/nar/gkg563.

Geluk A, Elferink DG, Slierendregt BL, Van Meijgaarden KE, Vries RRP, Ottenhoff THM, Bontrop RE. Evolutionary conservation of major histocompatibility complex-DR/peptide/T cell interactions in primates. Journal of Experimental Medicine. 1993; 177(4):979–987. doi: 10.1084/jem.177.4.979.

Gerstner C, Dubnovitsky A, Sandin C, Kozhukh G, Uchtenhagen H, James EA, Rönnelid J, Ytterberg AJ, Pieper J, Reed E, Tandre C, Rieck M, Zubarev RA, Rönnblom L, Sandalova T, Buckner JH, Achour A, Malmström V. Functional and structural characterization of a novel HLA-DRB1*04: 01-restricted α-enolase T cell epitope in rheumatoid arthritis. Frontiers in Immunology. 2016; 7:1–12. doi: 10.3389/fimmu.2016.00494.

Gleimer M, Wahl AR, Hickman HD, Abi-Rached L, Norman PJ, Guethlein LA, Hammond JA, Draghi M, Adams EJ, Juo S, Jalili R, Gharizadeh B, Ronaghi M, Garcia KC, Hildebrand WH, Parham P. Although Divergent in Residues of the Peptide Binding Site, Conserved Chimpanzee Patr-AL and Polymorphic Human HLA-A*02 Have Overlapping Peptide-Binding Repertoires. The Journal of Immunology. 2011; 186(3):1575–1588. doi: 10.4049/jimmunol.1002990.

885 **Grant EJ**, Josephs TM, Loh L, Clemens EB, Sant S, Bharadwaj M, Chen W, Rossjohn J, Gras S, Kedzierska K. Broad
886 CD8 + T cell cross-recognition of distinct influenza A strains in humans. Nature Communications. 2018;
887 9(1):1–16. http://dx.doi.org/10.1038/s41467-018-07815-5, doi: 10.1038/s41467-018-07815-5.

888 **Grant EJ**, Nguyen AT, Lobos CA, Szeto C, Chatzileontiadou DSM, Gras S. The unconventional role of HLA-E: The
889 road less traveled. Molecular Immunology. 2020; 120(December 2019):101–112. https://doi.org/10.1016/j.
890 molimm.2020.02.011, doi: 10.1016/j.molimm.2020.02.011.

891 **Grimholt U**, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM. A comprehensive analysis of teleost MHC
892 class I sequences. BMC Evolutionary Biology. 2015 12; 15(1):32. https://bmcevolbiol.biomedcentral.com/
893 articles/10.1186/s12862-015-0309-1, doi: 10.1186/s12862-015-0309-1.

894 **de Groot NG**, Otting N, Maccari G, Robinson J, Hammond JA, Blancher A, Lafont BAP, Guethlein LA, Wroblewski
895 EE, Marsh SGE, Shiina T, Walter L, Vigilant L, Parham P, O'Connor DH, Bontrop RE. Nomenclature report 2019:
896 major histocompatibility complex genes and alleles of Great and Small Ape and Old and New World monkey
897 species. Immunogenetics. 2020; 72(1-2):25–36. doi: 10.1007/s00251-019-01132-x.

898 **Guce AI**, Mortimer SE, Yoon T, Painter CA, Jiang W, Mellins ED, Stern LJ. HLA-DO acts as a substrate mimic to
899 inhibit HLA-DM by a competitive mechanism. Nature Structural and Molecular Biology. 2013; 20(1):90–98.
900 doi: 10.1038/nsmb.2460.

901 **Guethlein LA**, Norman PJ, Hilton HHG, Parham P. Co-evolution of MHC class I and variable NK cell receptors
902 in placental mammals. Immunological Reviews. 2015; 267(1):259–282. doi: 10.1111/imr.12326.

903 **Gyllensten UB**, Lashkari D, Erlich HA. Allelic diversification at the class II DQB locus of the mammalian major
904 histocompatibility complex. Proceedings of the National Academy of Sciences of the United States of America.
905 1990; 87(5):1835–1839. doi: 10.1073/pnas.87.5.1835.

906 **Hans JB**, Bergl RA, Vigilant L. Gorilla MHC class I gene and sequence variation in a comparative context. Im-
907 munogenetics. 2017; 69(5):303–323. doi: 10.1007/s00251-017-0974-x.

908 **Heijmans CMC**, de Groot NG, Bontrop RE. Comparative genetics of the major histocompatibility complex
909 in humans and nonhuman primates. International Journal of Immunogenetics. 2020; 47(3):243–260. doi:
910 10.1111/iji.12490.

911 **Henderson KN**, Tye-Din JA, Reid HH, Chen Z, Borg NA, Beissbarth T, Tatham A, Mannering SI, Purcell AW,
912 Dudek NL, van Heel DA, McCluskey J, Rossjohn J, Anderson RP. A Structural and Immunological Basis
913 for the Role of Human Leukocyte Antigen DQ8 in Celiac Disease. Immunity. 2007; 27(1):23–34. doi:
914 10.1016/j.immuni.2007.05.015.

915 **Hillig RC**, Hülsmeyer M, Saenger W, Welfle K, Misselwitz R, Welfle H, Kozerski C, Volz A, Uchanska-Ziegler B,
916 Ziegler A. Thermodynamic and Structural Analysis of Peptide- and Allele-dependent Properties of Two HLA-
917 B27 Subtypes Exhibiting Differential Disease Association. Journal of Biological Chemistry. 2004; 279(1):652–
918 663. doi: 10.1074/jbc.M307457200.

919 **Hinks A**, Bowes J, Cobb J, Ainsworth HC, Marion MC, Comeau ME, Sudman M, Han B, Becker ML, Bohnsack
920 JF, De Bakker PIW, Haas JP, Hazen M, Lovell DJ, Nigrovic PA, Nordal E, Punnaro M, Rosenberg AM, Rygg M,
921 Smith SL, et al. Fine-mapping the MHC locus in juvenile idiopathic arthritis (JIA) reveals genetic heterogeneity
922 corresponding to distinct adult inflammatory arthritic diseases. Annals of the Rheumatic Diseases. 2017;
923 76(4):765–772. doi: 10.1136/annrheumdis-2016-210025.

924 **Hirata J**, Hosomichi K, Sakaue S, Kanai M, Nakaoka H, Ishigaki K, Suzuki K, Akiyama M, Kishikawa T, Ogawa K,
925 Masuda T, Yamamoto K, Hirata M, Matsuda K, Momozawa Y, Inoue I, Kubo M, Kamatani Y, Okada Y. Genetic
926 and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. Nature
927 Genetics. 2019; 51(3):470–480. doi: 10.1038/s41588-018-0336-0.

928 **Hoare HL**, Sullivan LC, Pietra G, Clements CS, Lee EJ, Ely LK, Beddoe T, Falco M, Kjer-Nielsen L, Reid HH, Mc-
929 Cluskey J, Moretta L, Rossjohn J, Brooks AG. Structural basis for a major histocompatibility complex class
930 Ib-restricted T cell response. Nature Immunology. 2006; 7(3):256–264. doi: 10.1038/ni1312.

931 **Huan X**, Zhuo N, Lee HY, Ren EC. Allopurinol non-covalently facilitates binding of unconventional peptides
932 to HLA-B*58:01. Scientific Reports. 2023; 13(1):1–11. https://doi.org/10.1038/s41598-023-36293-z, doi:
933 10.1038/s41598-023-36293-z.

934 **Hughes AL**, Nei M. Nucleotide substitution at major histocompatibility complex class II loci: evidence for over-
935 dominant selection. Proceedings of the National Academy of Sciences of the United States of America. 1989;
936 86(3):958–962. doi: 10.1073/pnas.86.3.958.

Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveal overdominant selection. Nature. 1988; 335(6186):167–170. doi: 10.1038/335167a0.

Hurley CK. Naming HLA diversity: A review of HLA nomenclature. Human Immunology. 2021; 82(7):457–465. https://doi.org/10.1016/j.humimm.2020.03.005, doi: 10.1016/j.humimm.2020.03.005.

Igic B, Bohs L, Kohn JR. Ancient polymorphism reveals unidirectional breeding system shifts. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(5):1359–1363. doi: 10.1073/pnas.0506283103.

Illing PT, Pymm P, Croft NP, Hilton HG, Jojic V, Han AS, Mendoza JL, Mifsud NA, Dudek NL, McCluskey J, Parham P, Rossjohn J, Vivian JP, Purcell AW. HLA-B57 micropolymorphism defines the sequence and conformational breadth of the immunopeptidome. Nature Communications. 2018; 9(1). http://dx.doi.org/10.1038/s41467-018-07109-w, doi: 10.1038/s41467-018-07109-w.

Ioerger TR, Clark AG, Kao TH. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. Proceedings of the National Academy of Sciences of the United States of America. 1990; 87(24):9732–9735. doi: 10.1073/pnas.87.24.9732.

Jeffreys H. The Theory of Probability. 3 ed. OUP Oxford; 1998.

Jiang H, Wang CW, Wang Z, Dai Y, Zhu Y, Lee YS, Cao Y, Chung WH, Ouyang S, Wang H. Functional and structural characteristics of HLA-B*13:01-mediated specific T cells reaction in dapsone-induced drug hypersensitivity. Journal of Biomedical Science. 2022; 29(1):1–21. https://doi.org/10.1186/s12929-022-00845-8, doi: 10.1186/s12929-022-00845-8.

Jiang J, Taylor DK, Kim EJ, Boyd LF, Ahmad J, Mage MG, Truong HV, Woodward CH, Sgourakis NG, Cresswell P, Margulies DH, Natarajan K. Structural mechanism of tapasin-mediated MHC-I peptide loading in antigen presentation. Nature Communications. 2022; 13(1):1–13. doi: 10.1038/s41467-022-33153-8.

Jiang W, Birtley JR, Hung SC, Wang W, Chiou SH, Macaubas C, Kornum B, Tian L, Huang H, Adler L, Weaver G, Lu L, Ilstad-Minnihan A, Somasundaram S, Ayyangar S, Davis MM, Stern LJ, Mellins ED. In vivo clonal expansion and phenotypes of hypocretin-specific CD4+ T cells in narcolepsy patients and controls. Nature Communications. 2019; 10(1):1–17. http://dx.doi.org/10.1038/s41467-019-13234-x, doi: 10.1038/s41467-019-13234-x.

Kassardjian A, Sun E, Sookhoo J, Muthuraman K, Boligan KF, Kucharska I, Rujas E, Jetha A, Branch DR, Babiuk S, Barber B, Julien JP. Modular adjuvant-free pan-HLA-DR-immunotargeting subunit vaccine against SARS-CoV-2 elicits broad sarbecovirus-neutralizing antibody responses. Cell Reports. 2023; 42(4):112391. https://doi.org/10.1016/j.celrep.2023.112391, doi: 10.1016/j.celrep.2023.112391.

Kaufman J. The new W family reconstructs the evolution of MHC genes. Proceedings of the National Academy of Sciences. 2022; 119(5):119–121.

Kaur G, Gras S, Mobbs JI, Vivian JP, Cortes A, Barber T, Kuttikkatte SB, Jensen LT, Attfield KE, Dendrou CA, Carrington M, McVean G, Purcell AW, Rossjohn J, Fugger L. Structural and regulatory diversity shape HLA-C protein expression levels. Nature Communications. 2017 8; 8(1):15924. http://www.nature.com/articles/ncomms15924, doi: 10.1038/ncomms15924.

Kennedy AE, Ozbek U, Dorak MT. What has GWAS done for HLA and disease associations? International Journal of Immunogenetics. 2017; 44(5):195–211. doi: 10.1111/iji.12332.

Kiryu I, Dijkstra JM, Sarder RI, Fujiwara A, Yoshiura Y, Ototake M. New MHC class Ia domain lineages in rainbow trout (Oncorhynchus mykiss) which are shared with other fish species. Fish and Shellfish Immunology. 2005; 18(3):243–254. doi: 10.1016/j.fsi.2004.07.007.

Klein J. Generation of diversity at MHC loci: Implications for T-cell receptor repertoires. In: Fougereau M, Dausset J, editors. *Immunology*, vol. 80 London: Academic Press; 1980.p. 239–253.

Klein J. Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. Human Immunology. 1987; 19(3):155–162. doi: 10.1016/0198-8859(87)90066-8.

Kriener K, O'hUigin C, Tichy H, Klein J. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics. 2000 3; 51(3):169–178. http://link.springer.com/10.1007/s002510050028, doi: 10.1007/s002510050028.

Kriener K, O'hUigin C, Klein J. Independent origin of functional MHC class II genes in humans and new world monkeys. Human Immunology. 2001 1; 62(1):1–14. https://linkinghub.elsevier.com/retrieve/pii/S0198885900002330, doi: 10.1016/S0198-8859(00)00233-0.

Krishna C, Chiou J, Lee I, Kim HI, Aksit MA, Schack DV, Raychaudhuri S, Ziemek D, Hu X. The influence of HLA genetic variation on plasma protein expression. Nature Communications. 2024; 15(6469). http://dx.doi.org/10.1038/s41467-024-50583-8, doi: 10.1038/s41467-024-50583-8.

Kuderna LFK, Gao H, Janiak MC, Kuhlwilm M, Orkin JD, Bataillon T, Manu S, Valenzuela A, Bergman J, Rousselle M, Silva FE, Agueda L, Blanc J, Gut M, de Vries D, Goodhead I, Harris RA, Raveendran M, Jensen A, Chuma IS, et al. A global catalog of whole-genome diversity from 233 primate species. Science (New York, NY). 2023; 380(6648):906–913. doi: 10.1126/science.abn7829.

Kumar P, Vahedi-Faridi A, Saenger W, Merino E, López de Castro JA, Uchanska-Ziegler B, Ziegler A. Structural basis for T cell alloreactivity among three HLA-B14 and HLA-B27 antigens. Journal of Biological Chemistry. 2009; 284(43):29784–29797. doi: 10.1074/jbc.M109.038497.

Kundu S, Faulkes CG. A tangled history: Patterns of major histocompatibility complex evolution in the African mole-rats (Family: Bathyergidae). Biological Journal of the Linnean Society. 2007; 91(3):493–503. doi: 10.1111/j.1095-8312.2007.00814.x.

Kupfermann H, Mayer WE, O'hUigin C, Klein D, Klein J. Shared polymorphism between gorilla and human major histocompatibility complex DRB loci. Human Immunology. 1992 8; 34(4):267–278. https://linkinghub.elsevier.com/retrieve/pii/019888599290026J, doi: 10.1016/0198-8859(92)90026-J.

Kusano S, Kukimoto-Niino M, Satta Y, Ohsawa N, Uchikubo-Kamo T, Wakiyama M, Ikeda M, Terada T, Yamamoto K, Nishimura Y, Shirouzu M, Sasazuki T, Yokoyama S. Structural basis for the specific recognition of the major antigenic peptide from the Japanese cedar pollen allergen Cry j 1 by HLA-DP5. Journal of Molecular Biology. 2014; 426(17):3016–3027. doi: 10.1016/j.jmb.2014.06.020.

Lang HLE, Jacobsen H, Ikemizu S, Andersson C, Harlos K, Madsen L, Hjorth P, Sondergaard L, Svejgaard A, Wucherpfennig K, Stuart DI, Bell JI, Jones EY, Fugger L. A functional and structural basis for TCR cross-reactivity in multiple sclerosis. Nature Immunology. 2002; 3(10):940–943. doi: 10.1038/ni835.

Lawlor DA, Ward FE, Ennis PD, Jackson AP, Parham P. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature. 1988 9; 335(6187):268–271. http://www.nature.com/articles/335268a0, doi: 10.1038/335268a0.

Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, Noreen H, Oudshoorn M, Petersdorf E, Setterholm M, Spellman S, Weisdorf D, Williams TM, Anasetti C. High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation. Blood, The Journal of the American Society of Hematology. 2007; 110(13):4576–4583. doi: 10.1182/blood-2007-06-097386.

Leffler EM, Gao Z, Pfeifee S, Ségurel L, Auton A, Venn O, Bowden R, Bontrop R, Wall JD, Sella G, Donnelly P, McVean G, Przeworski M. Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. Science. 2013; 339(6127):1578–1582. doi: 10.1126/science.1234070.

Li L, Zhou X, Chen X. Characterization and Evolution of MHC Class II B Genes in Ardeid Birds. Journal of Molecular Evolution. 2011 6; 72(5-6):474–483. http://link.springer.com/10.1007/s00239-011-9446-3, doi: 10.1007/s00239-011-9446-3.

Li X, Singh NK, Collins DR, Ng R, Zhang A, Lamothe-Molina PA, Shahinian P, Xu S, Tan K, Piechocka-Trocha A, Urbach JM, Weber JK, Gaiha GD, Takou Mbah OC, Huynh T, Cheever S, Chen J, Birnbaum M, Zhou R, Walker BD, et al. Molecular basis of differential HLA class I-restricted T cell recognition of a highly networked HIV peptide. Nature Communications. 2023; 14(1). doi: 10.1038/s41467-023-38573-8.

Li Y, Li H, Martin R, Mariuzza RA. Structural basis for the binding of an immunodominant peptide from myelin basic protein in different registers by two HLA-DR2 proteins. Journal of Molecular Biology. 2000; 304(2):177–188. doi: 10.1006/jmbi.2000.4198.

Liu J, Chen KY, Ren EC. Structural insights into the binding of hepatitis B virus core peptide to HLA-A2 alleles: Towards designing better vaccines. European Journal of Immunology. 2011; 41(7):2097–2106. doi: 10.1002/eji.201041370.

Liu YC, Chen Z, Neller MA, Miles JJ, Purcell AW, McCluskey J, Burrows SR, Rossjohn J, Gras S. A molecular basis for the interplay between T cells, viral mutants, and human leukocyte antigen micropolymorphism. Journal of Biological Chemistry. 2014; 289(24):16688–16698. http://dx.doi.org/10.1074/jbc.M114.563502, doi: 10.1074/jbc.M114.563502.

Liu YC, Miles JJ, Neller MA, Gostick E, Price DA, Purcell AW, McCluskey J, Burrows SR, Rossjohn J, Gras S. Highly divergent T-cell receptor binding modes underlie specific recognition of a bulged viral peptide bound to a human leukocyte antigen class I molecule. Journal of Biological Chemistry. 2013; 288(22):15442–15454. http://dx.doi.org/10.1074/jbc.M112.447185, doi: 10.1074/jbc.M112.447185.

Loisel DA, Rockman MV, Wray GA, Altmann J, Alberts SC. Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(44):16331–16336. doi: 10.1073/pnas.0607662103.

Loll B, Rueckert C, Ziegler BU, Ziegler A, Crystal structure of HLA-B2709 complexed with the nona-peptide mA; 2020. https://doi.org/10.2210/pdb6Y27/pdb.

Lu D, Chen Y, Jiang M, Tan SG, Chai Y, Gao GF. Crystal structure of a TCR in complex with HLA-A*11:01 bound to KRAS peptide (VVGAVGVGK). PDB. 2023; .

Lugo JS, Cadavid LF. Patterns of MHC-G-like and MHC-B diversification in new world monkeys. PLoS ONE. 2015; 10(6):1–20. doi: 10.1371/journal.pone.0131343.

Luo Y, Kanai M, Choi W, Li X, Sakaue S, Yamamoto K, Ogawa K, Gutierrez-Arcelus M, Gregersen PK, Stuart PE, Elder JT, Forer L, Schönherr S, Fuchsberger C, Smith AV, Fellay J, Carrington M, Haas DW, Guo X, Palmer ND, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nature Genetics. 2021 10; 53(10):1504–1516. https://www.nature.com/articles/s41588-021-00935-7, doi: 10.1038/s41588-021-00935-7.

Maccari G, Robinson J, Ballingall K, Guethlein LA, Grimholt U, Kaufman J, Ho CS, de Groot NG, Flicek P, Bontrop RE, Hammond JA, Marsh SGE. IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. Nucleic Acids Research. 2017 1; 45(D1):D860–D864. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1050, doi: 10.1093/nar/gkw1050.

Maccari G, Robinson J, Hammond JA, Marsh SGE. The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. Immunogenetics. 2020 2; 72(1-2):49–55. http://link.springer.com/10.1007/s00251-019-01133-w, doi: 10.1007/s00251-019-01133-w.

Macdonald WA, Chen Z, Gras S, Archbold JK, Tynan FE, Clements CS, Bharadwaj M, Kjer-Nielsen L, Saunders PM, Wilce MCJ, Crawford F, Stadinsky B, Jackson D, Brooks AG, Purcell AW, Kappler JW, Burrows SR, Rossjohn J, McCluskey J. T Cell Allorecognition via Molecular Mimicry. Immunity. 2009; 31(6):897–908. http://dx.doi.org/10.1016/j.immuni.2009.09.025, doi: 10.1016/j.immuni.2009.09.025.

Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, Zhao M, Chennagiri N, Nordenfelt S, Tandon A, Skoglund P, Lazaridis I, Sankararaman S, Fu Q, Rohland N, Renaud G, Erlich Y, Willems T, Gallo C, Spence JP, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 2016 10; 538(7624):201–206. http://www.nature.com/articles/nature18964, doi: 10.1038/nature18964.

Marsh SGE, Albert ED, Bodmer WF, Bontrop RE, Dupont B, Erlich HA, Fernández-Viña M, Geraghty DE, Holdsworth R, Hurley CK, Lau M, Lee KW, MacH B, Maiers M, Mayr WR, Müller CR, Parham P, Petersdorf EW, Sasazuki T, Strominger JL, et al. Nomenclature for factors of the HLA system, 2010. Tissue Antigens. 2010; 75(4):291–455. doi: 10.1111/j.1399-0039.2010.01466.x.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, de Castro JMB, Carbonell E, Gerritsen F, et al. Genome-wide patterns of selection in 230 ancient Eurasians. Nature. 2015 12; 528(7583):499–503. http://www.nature.com/articles/nature16152, doi: 10.1038/nature16152.

Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G, Klein J. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. The EMBO journal. 1988; 7(9):2765–2774. doi: 10.1002/j.1460-2075.1988.tb03131.x.

Mayer W, O'hUigin C, Zaleska-Rutczynska Z, Klein J. Trans-species origin of Mhc-DRB polymorphism in the chimpanzee. Immunogenetics. 1992 10; 37(1):12–23. http://link.springer.com/10.1007/BF00223540, doi: 10.1007/BF00223540.

McConnell TJ, Talbot WS, McIndoe RA, Wakeland EK. The origin of MHC class II gene polymorphism in the genus Mus. Nature. 1988; 332(14):651–654. doi: 10.1111/j.1095-8312.1990.tb00826.x.

McKenzie LM, Pecon-Slattery J, Carrington M, O'Brien SJ. Taxonomic hierarchy of HLA class I allele sequences. Nature Genes and Immunity. 1999; 1(2):120–129. doi: 10.1038/sj.gene.6363648.

McLaren PJ, Ripke S, Pelak K, Weintrob AC, Patsopoulos NA, Jia X, Erlich RL, Lennon NJ, Kadie CM, Heckerman D, Gupta N, Haas DW, Deeks SG, Pereyra F, Walker BD, de Bakker PIW, Kuritzkes DR, Robbins GK, Shafer RW, Gulick RM, et al. Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in african americans. Human Molecular Genetics. 2012; 21(19):4334–4347. doi: 10.1093/hmg/dds226.

Miller JD, Weber DA, Ibegbu C, Pohl J, Altman JD, Jensen PE. Analysis of HLA-E Peptide-Binding Specificity and Contact Residues in Bound Peptide Required for Recognition by CD94/NKG2. The Journal of Immunology. 2003; 171(3):1369–1375. doi: 10.4049/jimmunol.171.3.1369.

Mobbs JI, Illing PT, Dudek NL, Brooks AG, Baker DG, Purcell AW, Rossjohn J, Vivian JP. The molecular basis for peptide repertoire selection in the human leucocyte antigen (HLA) C*06:02 molecule. Journal of Biological Chemistry. 2017; 292(42):17203–17215. doi: 10.1074/jbc.M117.806976.

Molineros JE, Looger LL, Kim K, Okada Y, Terao C, Sun C, Zhou Xj, Raj P, Kochi Y, Suzuki A, Akizuki S, Nakabo S, Bang SY, Lee HS, Kang YM, Suh CH, Chung WT, Park YB, Choe JY, Shim SC, et al. Amino acid signatures of HLA Class-I and II molecules are strongly associated with SLE susceptibility and autoantibody production in Eastern Asians. PLoS Genetics. 2019; 15(4):1–25. doi: 10.1371/journal.pgen.1008092.

Moradi S, Stankovic S, O'Connor GM, Pymm P, MacLachlan BJ, Faoro C, Retière C, Sullivan LC, Saunders PM, Widjaja J, Cox-Livingstone S, Rossjohn J, Brooks AG, Vivian JP. Structural plasticity of KIR2DL2 and KIR2DL3 enables altered docking geometries atop HLA-C. Nature Communications. 2021; 12(1). http://dx.doi.org/10.1038/s41467-021-22359-x, doi: 10.1038/s41467-021-22359-x.

Mosyak L, Zaller DM, Wiley DC. The structure of HLA-DM, the peptide exchange catalyst that loads antigen onto class II MHC molecules during antigen presentation. Immunity. 1998; 9(3):377–383. doi: 10.1016/S1074-7613(00)80620-2.

Motozono C, Kuse N, Sun X, Rizkallah PJ, Fuller A, Oka S, Cole DK, Sewell AK, Takiguchi M. Molecular Basis of a Dominant T Cell Response to an HIV Reverse Transcriptase 8-mer Epitope Presented by the Protective Allele HLA-B*51:01. The Journal of Immunology. 2014; 192(7):3428–3434. doi: 10.4049/jimmunol.1302667.

Müller NF, Bouckaert RR. Adaptive metropolis-coupled MCMC for BEAST 2. PeerJ. 2020; 8:1–16. doi: 10.7717/peerj.9473.

Naito T, Suzuki K, Hirata J, Kamatani Y, Matsuda K, Toda T, Okada Y. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. Nature Communications. 2021; 12(1):1–14. doi: 10.1038/s41467-021-21975-x.

Neefjes J, Jongsma MLM, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nature Reviews Immunology. 2011 12; 11(12):823–836. http://www.nature.com/articles/nri3084, doi: 10.1038/nri3084.

Nei M, Hughes AL. Polymorphism and evolution of the major histocompatibility complex loci in mammals. In: Selander R, Clark A, Whittam T, editors. *Evolution at the Molecular Level*, 6 ed. Sunderland, MA: Sinauer Associates, Inc; 1991.p. 222–247. doi: 10.1093/oxfordjournals.molbev.a040698.

Newman RM, Hall L, Connole M, Chen GL, Sato S, Yuste E, Diehl W, Hunter E, Kaur A, Miller GM, Johnson WE. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5α. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(50):19134–19139. doi: 10.1073/pnas.0605838103.

Nguyen TB, Jayaraman P, Bergseng E, Madhusudhan MS, Kim CY, Sollid LM. Unraveling the structural basis for the unusually rich association of human leukocyte antigen DQ2.5 with class-II-Associated invariant chain peptides. Journal of Biological Chemistry. 2017; 292(22):9218–9228. http://dx.doi.org/10.1074/jbc.M117.785139, doi: 10.1074/jbc.M117.785139.

Nicholson MJ, Moradi B, Seth NP, Xing X, Cuny GD, Stein RL, Wucherpfennig KW. Small Molecules That Enhance the Catalytic Efficiency of HLA-DM. The Journal of Immunology. 2006; 176(7):4208–4220. doi: 10.4049/jimmunol.176.7.4208.

Nielsen M, Lundegaard C, Blicher T, Lamberth K, Harndahl M, Justesen S, Røder G, Peters B, Sette A, Lund O, Buus S. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE. 2007; 2(8). doi: 10.1371/journal.pone.0000796.

Niu L, Cheng H, Zhang S, Tan S, Zhang Y, Qi J, Liu J, Gao GF. Structural basis for the differential classification of HLA-A*6802 and HLA-A*6801 into the A2 and A3 supertypes. Molecular Immunology. 2013; 55(3-4):381–392. http://dx.doi.org/10.1016/j.molimm.2013.03.015, doi: 10.1016/j.molimm.2013.03.015.

Nunes K, Maia MHT, dos Santos EJM, dos Santos SEB, Guerreiro JF, Petzl-Erler ML, Bedoya G, Gallo C, Poletti G, Llop E, Tsuneto L, Bortolini MC, Rothhammer F, Single R, Ruiz-Linares A, Rocha J, Meyer D. How natural selection shapes genetic differentiation in the MHC region: A case study with Native Americans. Human Immunology. 2021; 82(7):523–531. doi: 10.1016/j.humimm.2021.03.005.

Okada Y, Momozawa Y, Sakaue S, Kanai M, Ishigaki K, Akiyama M, Kishikawa T, Arai Y, Sasaki T, Kosaki K, Suematsu M, Matsuda K, Yamamoto K, Kubo M, Hirose N, Kamatani Y. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. Nature Communications. 2018 12; 9(1):1631. http://www.nature.com/articles/s41467-018-03274-0, doi: 10.1038/s41467-018-03274-0.

O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 2016 1; 44(D1):D733–D745. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1189, doi: 10.1093/nar/gkv1189.

Ooi JD, Petersen J, Tan YH, Huynh M, Willett ZJ, Ramarathinam SH, Eggenhuizen PJ, Loh KL, Watson KA, Gan PY, Alikhan MA, Dudek NL, Handel A, Hudson BG, Fugger L, Power DA, Holt SG, Coates PT, Gregersen JW, Purcell AW, et al. Dominant protection from HLA-linked autoimmunity by antigen-specific regulatory T cells. Nature. 2017; 545(7653):243–247. doi: 10.1038/nature22329.

Otting N, De Groot NG, Noort MC, Doxiadis GGM, Bontrop RE. Allelic diversity of Mhc-DRB alleles in rhesus macaques. Tissue Antigens. 2000 7; 56(1):58–68. http://doi.wiley.com/10.1034/j.1399-0039.2000.560108.x, doi: 10.1034/j.1399-0039.2000.560108.x.

Otting N, Kenter M, van Weeren P, Jonker M, Bontrop RE. Mhc-DQB repertoire variation in hominoid and Old World primate species. Journal of immunology (Baltimore, Md : 1950). 1992; 149(2):461–70. http://www.ncbi.nlm.nih.gov/pubmed/1624794.

Otting N, Bontrop RE. Evolution of the major histocompatibility complex DPA1 locus in primates. Human Immunology. 1995 2; 42(2):184–187. https://linkinghub.elsevier.com/retrieve/pii/0198885994000958, doi: 10.1016/0198-8859(94)00095-8.

Otting N, de Groot N, Doxiadis G, Bontrop R. Extensive Mhc-DQB variation in humans and non-human primate species. Immunogenetics. 2002 7; 54(4):230–239. http://link.springer.com/10.1007/s00251-002-0461-9, doi: 10.1007/s00251-002-0461-9.

Paganini J, Abi-Rached L, Gouret P, Pontarotti P, Chiaroni J, Di Cristofaro J. HLAIb worldwide genetic diversity: New HLA-H alleles and haplotype structure description. Molecular Immunology. 2019; 112(May):40–50. https://doi.org/10.1016/j.molimm.2019.04.017, doi: 10.1016/j.molimm.2019.04.017.

Parry CS, Gorski J, Stern LJ. Crystallographic Structure of the Human Leukocyte Antigen DRA, DRB3*0101: Models of a Directional Alloimmune Response and Autoimmunity. Journal of Molecular Biology. 2007; 371(2):435–446. doi: 10.1016/j.jmb.2007.05.025.

Petersen J, Kooy-Winkelaar Y, Loh KL, Tran M, van Bergen J, Koning F, Rossjohn J, Reid HH. Diverse T Cell Receptor Gene Usage in HLA-DQ8-Associated Celiac Disease Converges into a Consensus Binding Solution. Structure. 2016; 24(10):1643–1657. doi: 10.1016/j.str.2016.07.010.

Petrie EJ, Clements CS, Lin J, Sullivan LC, Johnson D, Huyton T, Heroux A, Hoare HL, Beddoe T, Reid HH, Wilce MCJ, Brooks AG, Rossjohn J. CD94-NKG2A recognition of human leukocyte antigen (HLA)-E bound to an HLA class I leader sequence. Journal of Experimental Medicine. 2008; 205(3):725–735. doi: 10.1084/jem.20072525.

Piontkivska H, Nei M. Birth-and-death evolution in primate MHC class I genes: Divergence time estimates. Molecular Biology and Evolution. 2003; 20(4):601–609. doi: 10.1093/molbev/msg064.

Pos W, Sethi DK, Call MJ, Schulze MSED, Anders AK, Pyrdol J, Wucherpfennig KW. Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. Cell. 2012; 151(7):1557–1568. http://dx.doi.org/10.1016/j.cell.2012.11.025, doi: 10.1016/j.cell.2012.11.025.

Purcell J, Lagunas-Robles G, Rabeling C, Borowiec ML, Brelsford A. The maintenance of polymorphism in an ancient social supergene. Molecular Ecology. 2021; 30(23):6246–6258. doi: 10.1111/mec.16196.

Racle J, Guillaume P, Schmidt J, Michaux J, Larabi A, Lau K, Perez MAS, Croce G, Genolet R, Coukos G, Zoete V, Pojer F, Bassani-Sternberg M, Harari A, Gfeller D. Machine learning predictions of MHC-II specificities reveal alternative binding mode of class II epitopes. Immunity. 2023; 56(6):1359–1375. doi: 10.1016/j.immuni.2023.03.009.

Radwan J, Babik W, Kaufman J, Lenz TL, Winternitz J. Advances in the Evolutionary Understanding of MHC Polymorphism. Trends in Genetics. 2020; 36(4):298–311. doi: 10.1016/j.tig.2020.01.008.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology. 2018 9; 67(5):901–904. https://academic.oup.com/sysbio/article/67/5/901/4989127, doi: 10.1093/sysbio/syy032.

Raychaudhuri S, Sandor C, Stahl EA, Freudenberg J, Lee HS, Jia X, Alfredsson L, Padyukov L, Klareskog L, Worthington J, Siminovitch KA, Bae SC, Plenge RM, Gregersen PK, De Bakker PIW. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nature Genetics. 2012; 44(3):291–296. doi: 10.1038/ng.1076.

Rist MJ, Theodossis A, Croft NP, Neller MA, Welland A, Chen Z, Sullivan LC, Burrows JM, Miles JJ, Brennan RM, Gras S, Khanna R, Brooks AG, McCluskey J, Purcell AW, Rossjohn J, Burrows SR. HLA Peptide Length Preferences Control CD8+ T Cell Responses. The Journal of Immunology. 2013; 191(2):561–571. doi: 10.4049/jimmunol.1300292.

Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. Nucleic Acids Research. 2019 10; 48(D1):D948–D955. https://academic.oup.com/nar/advance-article/doi/10.1093/nar/gkz950/5610347, doi: 10.1093/nar/gkz950.

Robinson J, Barker DJ, Marsh SGE. 25 years of the IPD-IMGT/HLA Database. Hla. 2024; 103(6):e15549. doi: 10.1111/tan.15549.

Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature. 2001 2; 409(6822):928–933. https://www.nature.com/articles/35057149, doi: 10.1038/35057149.

Sakaue S, Kanai M, Tanigawa Y, Karjalainen J, Kurki M, Koshiba S, Narita A, Konuma T, Yamamoto K, Akiyama M, Ishigaki K, Suzuki A, Suzuki K, Obara W, Yamaji K, Takahashi K, Asai S, Takahashi Y, Suzuki T, Shinozaki N, et al. A cross-population atlas of genetic associations for 220 human phenotypes. Nature Genetics. 2021; 53(10):1415–1424. doi: 10.1038/s41588-021-00931-x.

van de Sandt CE, Clemens EB, Grant EJ, Rowntree LC, Sant S, Halim H, Crowe J, Cheng AC, Kotsimbos TC, Richards M, Miller A, Tong SYC, Rossjohn J, Nguyen THO, Gras S, Chen W, Kedzierska K. Challenging immunodominance of influenza-specific CD8+ T cell responses restricted by the risk-associated HLA-A*68:01 allomorph. Nature Communications. 2019; 10(1). doi: 10.1038/s41467-019-13346-4.

Sano EB, Wall CA, Hutchins PR, Miller SR. Ancient balancing selection on heterocyst function in a cosmopolitan cyanobacterium. Nature Ecology & Evolution. 2018 3; 2(3):510–519. http://www.nature.com/articles/s41559-017-0435-9, doi: 10.1038/s41559-017-0435-9.

Satta Y, Mayer WE, Klein J. Evolutionary relationship of HLA-DRB genes inferred from intron sequences. Journal of Molecular Evolution. 1996 6; 42(6):648–657. http://link.springer.com/10.1007/BF02338798, doi: 10.1007/BF02338798.

Saunders PM, Vivian JP, Baschuk N, Beddoe T, Widjaja J, O'Connor GM, Hitchen C, Pymm P, Andrews DM, Gras S, McVicar DW, Rossjohn J, Brooks AG. The Interaction of KIR3DL1*001 with HLA Class I Molecules Is Dependent upon Molecular Microarchitecture within the Bw4 Epitope. The Journal of Immunology. 2015; 194(2):781–789. doi: 10.4049/jimmunol.1402542.

Sawyer SA, GENECONV: a computer package for the statistical detection of gene conversion. Distributed by the author, Department of Mathematics, Washington University in St. Louis; 1999. http://www.math.wustl.edu/~sawyer.

Scally SW, Law SC, Ting YT, Heemst JV, Sokolove J, Deutsch AJ, Bridie Clemens E, Moustakas AK, Papadopoulos GK, Woude DVD, Smolik I, Hitchon CA, Robinson DB, Ferucci ED, Bernstein CN, Meng X, Anaparti V, Huizinga T, Kedzierska K, Reid HH, et al. Molecular basis for increased susceptibility of Indigenous North Americans to seropositive rheumatoid arthritis. Annals of the Rheumatic Diseases. 2017 11; 76(11):1915–1923. https://ard.bmj.com/lookup/doi/10.1136/annrheumdis-2017-211300, doi: 10.1136/annrheumdis-2017-211300.

Schulze MSED, Wucherpfennig KW. The mechanism of HLA-DM induced peptide exchange in the MHC class II antigen presentation pathway. Current Opinion in Immunology. 2012; 24(1):105–111. doi: 10.1016/j.coi.2011.11.004.

Schutte RJ, Li D, Ostrov DA, HLA-B*15:01 complexed with a synthetic peptide; 2020. https://doi.org/10.2210/pdb6UZP/pdb.

Sebastian Klobuch, Lim JJ, van Balen P, Kester MGD, Klerk Wd, Ru AHd, Pothast CR, Jedema I, Drijfhout JW, Rossjohn J, Reid HH, Veelen PAv, Falkenburg JHF, Heemskerk MHM. Human T cells recognize HLA-DP–bound peptides in two orientations. Proceedings of the National Academy of Sciences. 2022; 119(49):e2214331119. doi: 10.1073/pnas.

Ségurel L, Thompson EE, Flutre T, Lovstad J, Venkat A, Margulis SW, Moyse J, Ross S, Gamble K, Sella G, Ober C, Przeworski M. The ABO blood group is a trans-species polymorphism in primates. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(45):18493–18498. doi: 10.1073/pnas.1210603109.

Sharon E, Sibener LV, Battle A, Fraser HB, Christopher Garcia K, Pritchard JK. Genetic variation in MHC proteins is associated with T cell receptor expression biases. Nature Genetics. 2016; 48(9):995–1002. doi: 10.1038/ng.3625.

Shiroishi M, Kuroki K, Rasubala L, Tsumoto K, Kumagai I, Kurimoto E, Kato K, Kohda D, Maenaka K. Structural basis for recognition of the nonclassical MHC molecule HLA-G by the leukocyte Ig-like receptor B2 (LILRB2/LIR2/ILT4/CD85d). Proceedings of the National Academy of Sciences of the United States of America. 2006; 103(44):16412–16417. doi: 10.1073/pnas.0605228103.

Sian TCCLK, Indumathy S, Halim H, Greule A, Cryle MJ, Bowness P, Rossjohn J, Gras S, Purcell AW, Schittenhelm RB. Allelic association with ankylosing spondylitis fails to correlate with human leukocyte antigen b27 homodimer formation. Journal of Biological Chemistry. 2019; 294(52):20185–20195. http://dx.doi.org/10.1074/jbc.RA119.010257, doi: 10.1074/jbc.RA119.010257.

Simons ND, Eick GN, Ruiz-Lopez MJ, Omeja PA, Chapman CA, Goldberg TL, Ting N, Sterner KN. Cis-regulatory evolution in a wild primate: Infection-associated genetic variation drives differential expression of MHC-DQA1 in vitro. Molecular Ecology. 2017; 26(17):4523–4535. doi: 10.1111/mec.14221.

Slierendregt BL, van Noort JT, Bakas RM, Otting N, Jonker M, Bontrop RE. Evolutionary stability of transspecies major histocompatibility complex class II DRB lineages in humans and rhesus monkeys. Human Immunology. 1992 9; 35(1):29–39. https://linkinghub.elsevier.com/retrieve/pii/0198885992900922, doi: 10.1016/0198-8859(92)90092-2.

Slierendregt BL, Otting N, Kenter M, Bontrop RE. Allelic diversity at the Mhc-DP locus in rhesus macaques (Macaca mulatta). Immunogenetics. 1995 1; 41(1):29–37. http://link.springer.com/10.1007/BF00188429, doi: 10.1007/BF00188429.

Smith CJ, Strausz S, Spence JP, Ollila HM, Pritchard JK. Haplotype Analysis Reveals Pleiotropic Disease Associations in the HLA Region. medRxiv. 2024; doi: https://doi.org/10.1101/2024.07.29.24311183.

Smith KJ, Pyrdol J, Gauthier L, Wiley DC, Wucherpfennig KW. Crystal structure of HLA-DR2 (DRA*0101, DRB1*1501) Complexed with a peptide from human myelin basic protein. Journal of Experimental Medicine. 1998; 188(8):1511–1520. doi: 10.1084/jem.188.8.1511.

Stuart PE, Tsoi LC, Nair RP, Ghosh M, Kabra M, Shaiq PA, Raja GK, Qamar R, Thelma BK, Patrick MT, Parihar A, Singh S, Khandpur S, Kumar U, Wittig M, Degenhardt F, Tejasvi T, Voorhees JJ, Weidinger S, Franke A, et al. Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide. Human Genetics and Genomics Advances. 2022; 3(1):100069. doi: 10.1016/j.xhgg.2021.100069.

Sullivan LC, Walpole NG, Farenc C, Pietra G, Sum MJW, Clements CS, Lee EJ, Beddoe T, Falco M, Mingari MC, Moretta L, Gras S, Rossjohn J, Brooks AG. A conserved energetic footprint underpins recognition of human leukocyte antigen-E by two distinct $\alpha\beta$ T cell receptors. Journal of Biological Chemistry. 2017; 292(51):21149–21158. http://dx.doi.org/10.1074/jbc.M117.807719, doi: 10.1074/jbc.M117.807719.

Sun M, Liu J, Qi J, Tefsen B, Shi Y, Yan J, Gao GF. Nα-Terminal Acetylation for T Cell Recognition: Molecular Basis of MHC Class I–Restricted Nα-Acetylpeptide Presentation. The Journal of Immunology. 2014; 192(12):5509–5519. doi: 10.4049/jimmunol.1400199.

Teixeira JC, de Filippo C, Weihmann A, Meneu JR, Racimo F, Dannemann M, Nickel B, Fischer A, Halbwax M, Andre C, Atencia R, Meyer M, Parra G, Pääbo S, Andrés AM. Long-Term Balancing Selection in LAD1 Maintains a Missense Trans-Species Polymorphism in Humans, Chimpanzees, and Bonobos. Molecular Biology and Evolution. 2015 5; 32(5):1186–1196. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/msv007, doi: 10.1093/molbev/msv007.

Teze D, Hendrickx J, Czjzek M, Ropartz D, Sanejouand YH, Tran V, Tellier C, Dion M. Semi-rational approach for converting a GH1 β-glycosidase into a β-transglycosidase. Protein Engineering, Design and Selection. 2014; 27(1):13–19. doi: 10.1093/protein/gzt057.

Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, Hinds DA. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nature Communications. 2017; 8(1). doi: 10.1038/s41467-017-00257-5.

Ting YT, Dahal-Koirala S, Kim HSK, Qiao SW, Neumann RS, Lundin KEA, Petersen J, Reid HH, Sollid LM, Rossjohn J. A molecular basis for the T cell response in HLA-DQ2.2 mediated celiac disease. Proceedings of the National Academy of Sciences of the United States of America. 2020; 117(6):3063–3073. doi: 10.1073/pnas.1914308117.

Tollefsen S, Hotta K, Chen X, Simonsen B, Swaminathan K, Mathews II, Sollid LM, Kim CY. Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. Journal of Biological Chemistry. 2012; 287(17):13611–13619. doi: 10.1074/jbc.M111.320374.

Vivian JP, Rossjohn J, HLA-B*27:05 in complex with the pan-HLA-Ia monoclonal antibody W6/32; 2022.

Vizcaíno JA, Kubiniok P, Kovalchik KA, Ma Q, Duquette JD, Mongrain I, Deutsch EW, Peters B, Sette A, Sirois I, Caron E. The human immunopeptidome project: A roadmap to predict and treat immune diseases. Molecular and Cellular Proteomics. 2020; 19(1):31–49. doi: 10.1074/mcp.R119.001743.

Vollmers S, Lobermeyer A, Körner C. The new kid on the block: Hla-c, a key regulator of natural killer cells in viral immunity. Cells. 2021; 10(11). doi: 10.3390/cells10113108.

Waage J, Standl M, Curtin JA, Jessen LE, Thorsen J, Tian C, Schoettler N, Flores C, Abdellaoui A, Ahluwalia TS, Alves AC, Amaral AFS, Antó JM, Arnold A, Barreto-Luis A, Baurecht H, van Beijsterveldt CEM, Bleecker ER, Bonàs-Guarch S, Boomsma DI, et al. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. Nature Genetics. 2018; 50(8):1072–1080. doi: 10.1038/s41588-018-0157-1.

Wakeland EK, Tarruzzee RW, Lu CC, Potts W, McIndoe RA, Talbot WS, McConnell TJ. The evolution of MHC class II genes within the genus Mus. In: David CS, editor. *H-2 Antigens: Genes, Molecules, Function*, 1 ed. Boston, MA: Springer; 1987.p. 139–153. doi: 10.1007/978-1-4757-0764-9{_}14.

Walpole NG, Kjer-Nielsen L, Kostenko L, McCluskey J, Brooks AG, Rossjohn J, Clements CS. The Structure and Stability of the Monomorphic HLA-G Are Influenced by the Nature of the Bound Peptide. Journal of Molecular Biology. 2010; 397(2):467–480. http://dx.doi.org/10.1016/j.jmb.2010.01.052, doi: 10.1016/j.jmb.2010.01.052.

Walters LC, Rozbesky D, Harlos K, Quastel M, Sun H, Springer S, Rambo RP, Mohammed F, Jones EY, McMichael AJ, Gillespie GM. Primary and secondary functions of HLA-E are determined by stability and conformation of the peptide-bound complexes. Cell Reports. 2022; 39(11):110959. https://doi.org/10.1016/j.celrep.2022.110959, doi: 10.1016/j.celrep.2022.110959.

Wroblewski EE, Guethlein LA, Norman PJ, Li Y, Shaw CM, Han AS, Ndjango JBN, Ahuka-Mundeke S, Georgiev AV, Peeters M, Hahn BH, Parham P. Bonobos Maintain Immune System Diversity with Three Functional Types of MHC-B. The Journal of Immunology. 2017; 198(9):3480–3493. doi: 10.4049/jimmunol.1601955.

Wu CH, Suchard MA, Drummond AJ. Bayesian Selection of Nucleotide Substitution Models and Their Site Assignments. Molecular Biology and Evolution. 2013 3; 30(3):669–688. https://academic.oup.com/mbe/article-lookup/doi/10.1093/molbev/mss258, doi: 10.1093/molbev/mss258.

Wu Y, Gao F, Liu J, Qi J, Gostick E, Price DA, Gao GF. Structural Basis of Diverse Peptide Accommodation by the Rhesus Macaque MHC Class I Molecule Mamu-B*17: Insights into Immune Protection from Simian Immunodeficiency Virus. The Journal of Immunology. 2011; 187(12):6382–6392. doi: 10.4049/jimmunol.1101726.

Xu SX, Ren WH, Li SZ, Wei FW, Zhou KY, Yang G. Sequence Polymorphism and Evolution of Three Cetacean MHC Genes. Journal of Molecular Evolution. 2009 9; 69(3):260–275. http://link.springer.com/10.1007/s00239-009-9272-z, doi: 10.1007/s00239-009-9272-z.

Yagita Y, Kuse N, Kuroki K, Gatanaga H, Carlson JM, Chikata T, Brumme ZL, Murakoshi H, Akahoshi T, Pfeifer N, Mallal S, John M, Ose T, Matsubara H, Kanda R, Fukunaga Y, Honda K, Kawashima Y, Ariumi Y, Oka S, et al. Distinct HIV-1 Escape Patterns Selected by Cytotoxic T Cells with Identical Epitope Specificity. Journal of Virology. 2013; 87(4):2253–2263. doi: 10.1128/jvi.02572-12.

Yamamoto Y, Morita D, Shima Y, Midorikawa A, Mizutani T, Suzuki J, Mori N, Shiina T, Inoko H, Tanaka Y, Mikami B, Sugita M. Identification and Structure of an MHC Class I–Encoded Protein with the Potential to Present N -Myristoylated 4-mer Peptides to T Cells . The Journal of Immunology. 2019; 202(12):3349–3358. doi: 10.4049/jimmunol.1900087.

Yasumizu Y, Sakaue S, Konuma T, Suzuki K, Matsuda K, Murakami Y, Kubo M, Palamara PF, Kamatani Y, Okada Y. Genome-Wide Natural Selection Signatures Are Linked to Genetic Risk of Modern Phenotypes in the Japanese Population. Molecular Biology and Evolution. 2020 5; 37(5):1306–1316. https://academic.oup.com/mbe/article/37/5/1306/5709748, doi: 10.1093/molbev/msaa005.

Yin L, Crawford F, Marrack P, Kappler JW, Dai S. T-cell receptor (TCR) interaction with peptides that mimic nickel offers insight into nickel contact allergy. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(45):18517–18522. doi: 10.1073/pnas.1215928109.

Zavala-Ruiz Z, Strug I, Anderson MW, Gorski J, Stern LJ. A Polymorphic Pocket at the P10 Position Contributes to Peptide Binding Specificity in Class II MHC Proteins. Chemistry & Biology. 2004; 11:1395–1402. doi: 10.1016/j.

Zhang S, Liu J, Cheng H, Tan S, Qi J, Yan J, Gao GF. Structural basis of cross-allele presentation by HLA-A*0301 and HLA-A*1101 revealed by two HIV-derived peptide complexes. Molecular Immunology. 2011; 49(1-2):395–401. http://dx.doi.org/10.1016/j.molimm.2011.08.015, doi: 10.1016/j.molimm.2011.08.015.

Zhao M, Wang Y, Shen H, Li C, Chen C, Luo Z, Wu H. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. BMC Evolutionary Biology. 2013 12; 13(1):113. https://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-13-113, doi: 10.1186/1471-2148-13-113.

Zhou F, Cao H, Zuo X, Zhang T, Zhang X, Liu X, Xu R, Chen G, Zhang Y, Zheng X, Jin X, Gao J, Mei J, Sheng Y, Li Q, Liang B, Shen J, Shen C, Jiang H, Zhu C, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. Nature Genetics. 2016; 48(7):740–746. doi: 10.1038/ng.3576.

Zhu S, Liu K, Chai Y, Wu Y, Lu D, Xiao W, Cheng H, Zhao Y, Ding C, Lyu J, Lou Y, Gao GF, Liu WJ. Divergent Peptide Presentations of HLA-A*30 Alleles Revealed by Structures With Pathogen Peptides. Frontiers in immunology. 2019; 10:1709. doi: 10.3389/fimmu.2019.01709.