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

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

[1]Department of Biology, Stanford University, Stanford, CA USA
[2]Department of Genetics, Stanford University, Stanford, CA USA
Email: afortier@stanford.edu, pritch@stanford.edu

June 29, 2022

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. 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 used Bayesian phylogenetic methods to determine the extent of TSP at six classical MHC genes. We find strong support for TSP in all six genes, including between humans and old-world monkeys in HLA-DRB1 and even—remarkably—between humans and new-world monkeys in HLA-DQB1. Despite the long-term persistence of ancient lineages, we additionally observe rapid evolution at amino acids within the peptide-binding domain. The most rapidly-evolving positions are also strongly enriched for autoimmune and infectious disease associations. Together, these results suggest complex selective forces arising from differential peptide binding, which drive short-term allelic turnover within lineages while also maintaining deeply divergent lineages for at least 45 million years.

*ORCID ID: 0000-0001-5964-2540 (A.L.F.)

Introduction

The Major Histocompatibility Complex (MHC) is a large immunity locus shared among the jawed vertebrates (Figure 1A)[1]. In humans, the MHC is also known as the HLA (Human Leukocyte Antigen) region; it spans about 5 MB on Chromosome 6 and contains 412 total genes[2,3]. This includes the "classical" MHC genes that 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[4].

The MHC locus is extraordinarily polymorphic, with thousands of distinct alleles (i.e., haplotypes) observed at the classical genes[5–7]. Different alleles are functionally diverse, with distinct peptide-binding affinities and, consequently, allelic differences in pathogen detection[4]. 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[8]. Here, we aim to understand the evolution of the MHC by characterizing trans-species polymorphism in six classical genes, HLA-A, -B, -C, -DPB1, -DQB1 and -DRB1, and exploring the functional consequences.

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[9,10], alleles shared across species[11,12], and high nucleotide diversity across the region[13,14] in rodents and primates. Indeed, modern data show that nucleotide diversity in the human MHC exceeds 70-times the genome-wide average near the classical genes, suggesting ancient balancing selection (Figure 1A-C). Meanwhile, the MHC also features prominently in genome-wide scans for short-term directional selection[15–20].

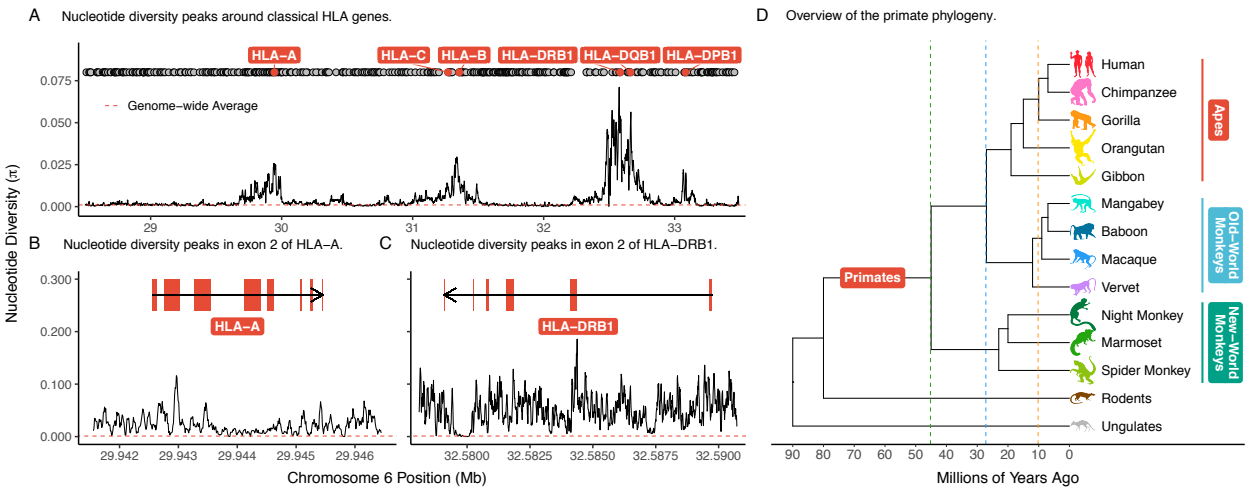

Figure 1: **The classical MHC region. A)** Each point at top represents the location of a gene, including the Class I HLA genes A, C, and B, and Class II HLA genes DRB1, DQB1, and DPB1 that are the focus of this paper. The black line shows nucleotide diversity (Nei and Li's π) across the region, while the red dotted line shows the genomewide average nucleotide diversity ($\pi \approx 0.001$)[21]. **B)** Nucleotide diversity around HLA-A, with exon structure shown. **C)** Nucleotide diversity around HLA-DRB1, with exon structure shown. Zoomed-in views of nucleotide diversity for the other genes is shown in Supplementary Figure S1. **D)** Species tree showing the phylogenetic relationships among selected primates from this study. The orange dashed line indicates the split of the human/chimpanzee lineage from gorilla, approximately 10 million years ago (Mya). The blue dashed line indicates the split of the ape lineage from OWM, approximately 27 Mya. The green line indicates the split of the Catarrhine lineage (apes and OWM) from NWM, approximately 45 Mya. References for divergence times are listed in Supplementary Table S1.

Selection at the MHC is believed to reflect the dynamics of host-pathogen evolution[1,22]. The MHC system

provides constant surveillance against a diverse and ever-changing array of pathogens, all while avoiding detection of self-peptides, which can lead to autoimmune disease. However, the precise mode of selection has not been fully resolved, and several models can explain salient features of the MHC. Under *heterozygote advantage*, individuals with different MHC alleles at each classical locus could defend against a wider range of pathogens than homozygotes. This would drive selection in favor of new alleles (as these would nearly always be heterozygous), and select against common alleles. The *divergent allele advantage* model extends the heterozygote advantage model by noting that heterozygotes with functionally divergent alleles may have better immune surveillance than heterozygotes with similar alleles[1,23]. Other models emphasize the role of pathogen evolution. In the *rare allele advantage/frequency dependent selection* models, pathogens adapt to avoid detection by the most frequent MHC alleles, meaning novel alleles enjoy fitness advantages[1,24]. Lastly, in *fluctuating selection*, pathogen pressures vary across time and space. Specific MHC alleles could become advantageous during epidemics, resulting in frequency shifts over time or across geographical areas[1]. In all of these modes, high diversity and long-term maintenance of alleles are inextricably linked.

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[25]. Indeed, only ∼100 loci genome-wide show compelling evidence for sharing of ancestral alleles between humans and our closest relatives, chimpanzees[26]. TSP among humans and more distantly-related species is even rarer; besides MHC, the only other clear example of deep TSP is at the ABO locus[27]. 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 ∼20 million years ago[28]. 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 was first proposed in the 1980s on the basis of unusual sequence similarity between mice and rats[11,13,29–32], and between humans and chimpanzees[12,33]. Later work has reported likely TSP between humans and apes[34–37] and humans and old world monkeys[38–48,48–50] (see Supplement for more details); 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 depth at each gene is still uncertain.

Thus, despite the long history of evolutionary studies of the MHC, we still lack a full description of key aspects of variation in the classical MHC genes. In particular, while the high within-species diversity and between-species allele sharing have been well-known for 30 years, we still do not know precisely how ancient these lineages are, how this varies across genes, how to interpret the evidence for ancient lineages alongside the evidence for rapid evolution at specific sites, or how these evolutionary features relate to protein function and disease associations.

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[5–7]. The thousands of full-length allele sequences deposited here provide a snapshot of MHC diversity across a wide range of species and allow us to answer questions at a larger scale than previously (Figure 1D). We account for the uncertainty in phylogenetic inference using a Bayesian MCMC approach (*BEAST2*) that samples from the posterior distribution of trees, allowing us to perform formal model-testing for shared ancestral lineages. *BEAST2* also permits evolutionary rates to vary across sites, thus allowing us to minimize the influence of rapidly-evolving, potentially-convergent positions. Lastly, we integrated this analysis with protein structural information and disease associations.

We find conclusive support for TSP among the great apes for HLA-B and -DPB1, and between humans and OWM in HLA-DRB1. Remarkably, we observe TSP in HLA-DQB1 going back to the common ancestor of humans and NWM, implying that alleles have been maintained by balancing selection for at least 45 million years. We also find that most rapidly-evolving sites in all six classical genes are located in the critical peptide-binding regions; moreover, the most rapidly-evolving sites are also the most 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 these genes, and motivate further evolutionary and functional studies to better understand this unique system.

Results

Data. We collated sequence data for six classical MHC genes from existing databases. The IPD-IMGT/HLA database contains over 32,800 human alleles[51], while the IPD-MHC database contains over 11,400 alleles across 77 species, including 55 non-human primates[52]. Most alleles belong to the highly polymorphic classical MHC genes, and most are partial sequences covering exons 2 and 3 (which contain the peptide binding region), while the rest span the entire coding regions of genes. A handful of alleles also include the introns, although such full sequences are limited to only a few species. We compiled alleles of six highly polymorphic classical genes, HLA-A, -B, -C, -DPB1, -DQB1, and -DRB1, for which the entire coding sequence (all exons) was available in the database (details in Methods). Because phylogenetic inference is computationally intensive, we reduced each set to a representative collection of alleles spanning the species' allelic diversity, according to the alleles' two-digit designations. The final data set consisted of 149 alleles for HLA-A, 197 for HLA-B, 61 for HLA-C, 100 alleles for HLA-DPB1, 122 alleles for HLA-DQB1, and 122 alleles for HLA-DRB1 (available as Supplementary Files).

Phylogenetic Inference. To assess the evidence for ancient TSP, we first reconstructed the phylogenetic relationships between alleles. This task is potentially challenging for several reasons: (i) phylogenetic inference is subject to considerable uncertainty, and it can be difficult to quantify (and visualize) the uncertainty in tree reconstruction; (ii) adaptive evolution of missense variants in the peptide-binding regions could make divergent alleles appear similar due to convergent evolution; (iii) recombination means that different regions of a gene have different ancestral histories, thus not conforming to standard phylogenetic assumptions.

To address these challenges, we used a method for Bayesian evolutionary analysis, *BEAST2*, to estimate allele trees[53,54]. Previous MHC work has mostly relied on either summary trees or rudimentary assessments of tree confidence such as bootstrap clade support, which do not naturally allow for hypothesis testing. As an MCMC method, *BEAST2* produces a large set of phylogenies, drawn from the posterior distribution, that we could use for formal model-testing. Second, to account for rapidly-evolving sites, we allowed evolutionary rates to vary among nucleotide sites. The package *SubstBMA*[55] allows *BEAST2* to estimate partitions among sites with different evolutionary rates for each partition, thereby preventing rapidly-evolving sites from disproportionately shaping the final phylogeny. The Bayesian framework also allowed us to incorporate priors, average over models, and quantify uncertainty. Third, to minimize the impact of recombination, we mainly describe the analysis of short regions—a single exon in most cases—even though this may increase uncertainty due to the smaller numbers of sites. We note that, if anything, recombination should bias the data *away* from observing TSP since recombination homogenizes haplotypes within species but not between. Nonetheless, we additionally support our main conclusions with an analysis of synonymous divergence distances that is independent of the phylogenetic analyses.

Trans-Species Polymorphism is Widespread. We considered six classical genes and five different subsets of each gene: exon 2 alone, exon 3 alone, exons 2 and 3 together (PBR), the non-PBR exons to-

gether (Other Exons), and all exons together (CDS). Because few intron sequences were available for non-human species, we did not include them in our analyses. For all six genes, the summary trees suggest deep sharing of ancient lineages among species.

Two of these trees are shown in Figure 2, corresponding to the second exons of HLA-DRB1 and HLA-DQB1, respectively (the other exons and genes are in Supplementary Figures S2-S7). Each image shows a single phylogeny that maximizes the product of posterior clade probabilities[56]. Each tip represents an IPD-MHC/HLA allele, named according to the standard hierarchical naming scheme[57] and with color and shape indicating the species. Clade labels were chosen somewhat arbitrarily to make it easier to indicate key interpretative features of the trees. Outgroup and single-species clades are collapsed for clarity.

Critically, we observe that, at both genes, the alleles fail to cluster together according to species, as indicated by the mixed-color clades throughout the trees. At HLA-DRB1, the human alleles (in red) are

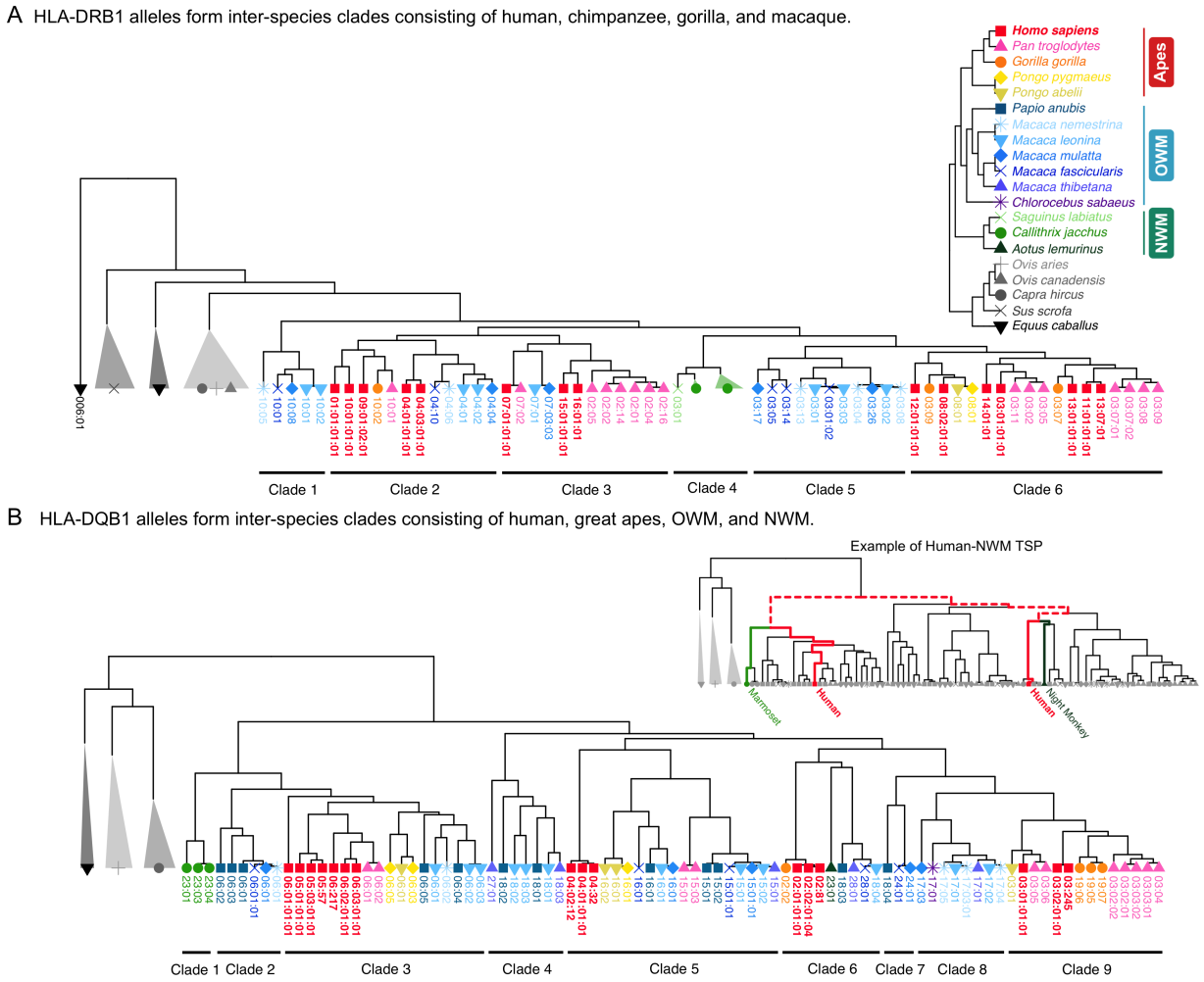

Figure 2: **BEAST2 allele summary trees using sequences from exon 2. A)** HLA-DRB1 and **B)** HLA-DQB1. Each tip represents an allele, with color and shape representing the species. Human alleles (red squares) are also bolded for emphasis. Outgroup and single-species clades are collapsed for clarity. The color/shape key (upper right) also depicts the species tree. The smaller inset tree in panel B highlights the relationships between two human alleles (red) and two NWM alleles (green). The indicated human and NWM lineages coalesce more recently between groups than within each group. This particular example of ancient TSP has extremely strong statistical support (Bayes factor > 7000).

spread across the tree, appearing in Clades 2, 3, and 6. Chimpanzee alleles (pink) also appear in Clades 2, 3, and 6. Even macaque alleles (blue) are distributed among several of the same clades: 2, 3, and 4. Hence, alleles are often most closely related to alleles from other species, suggesting extensive TSP and sharing of ancient lineages at least as far back as the ape–old-world monkey split.

The HLA-DQB1 tree suggests even more extensive TSP: human and other great ape alleles are found in Clades 3, 4, 6, and 9; old-world monkeys are found in Clades 2, 3, 4, 5, 6, 7, and 8; new-world monkeys are in Clades 1 and 6. The inset figure highlights an example of TSP between human and new-world monkey alleles. This particular example has strong statistical support. Thus, remarkably, data from HLA-DQB1 indicate the persistence of distinct ancestral lineages for at least 45 million years.

However, while the 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 (and respectively from B) are closest in the unrooted tree, this quartet supports monophyly of A (and of B). But if alleles from A and B are closest 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. The precise interpretation of Bayes factors depends on one's prior expectation; however, following standard guidelines[58] 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.

Bayes factors are shown in Figure 3. For each genic region (x-axis), we tested for TSP among humans, non-human apes, OWM, and NWM, 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. Red indicates support for TSP among the species in that category, while blue indicates evidence against TSP.

At the Class I genes, HLA-A, -B, and -C, we find strong support for TSP within the great apes: between human, chimpanzee, and gorilla for HLA-A and -C, and also with orangutan at -B. We do not find evidence for deeper TSP within Class I (but note that we do not have gibbon alleles apart from HLA-A). Meanwhile, the results for HLA-DRB1 and -DQB1 confirm the presence of ancient TSP. Bayes factors for exon 2 of HLA-DRB1 indicate TSP between humans, chimpanzees, gorillas, and even OWM. (The Bayes factors between human and orangutan are 0, but there are only two orangutan alleles in this dataset.) Even more striking, the Bayes factors indicate TSP in exon 2 of HLA-DQB1 at least as far back as our ancestors with NWM.

There is also extensive support for TSP within the OWM for all genes (except HLA-C, which is only present in the great apes), and within NWM for HLA-B and -DPB1. HLA-DQB1 and -DPB1 also show TSP between OWM and NWM. There is no support for TSP between humans and any of the outgroup species. Supplementary Figures S9-S15 show Bayes factors for all analyses. Thus, in summary, the phylogenetic analyses strongly support the presence of ancient TSP in these genes.

However, the tree-based model has certain limitations: it may not fully account for convergent evolution at functional sites (even though we allow for rate variation among sites) and it ignores recombination. Therefore, we next used a complementary, nonparametric approach based on synonymous nucleotide distances (dS) between pairs of alleles to verify the plausibility of our tree-based results. Pairwise synonymous dis-

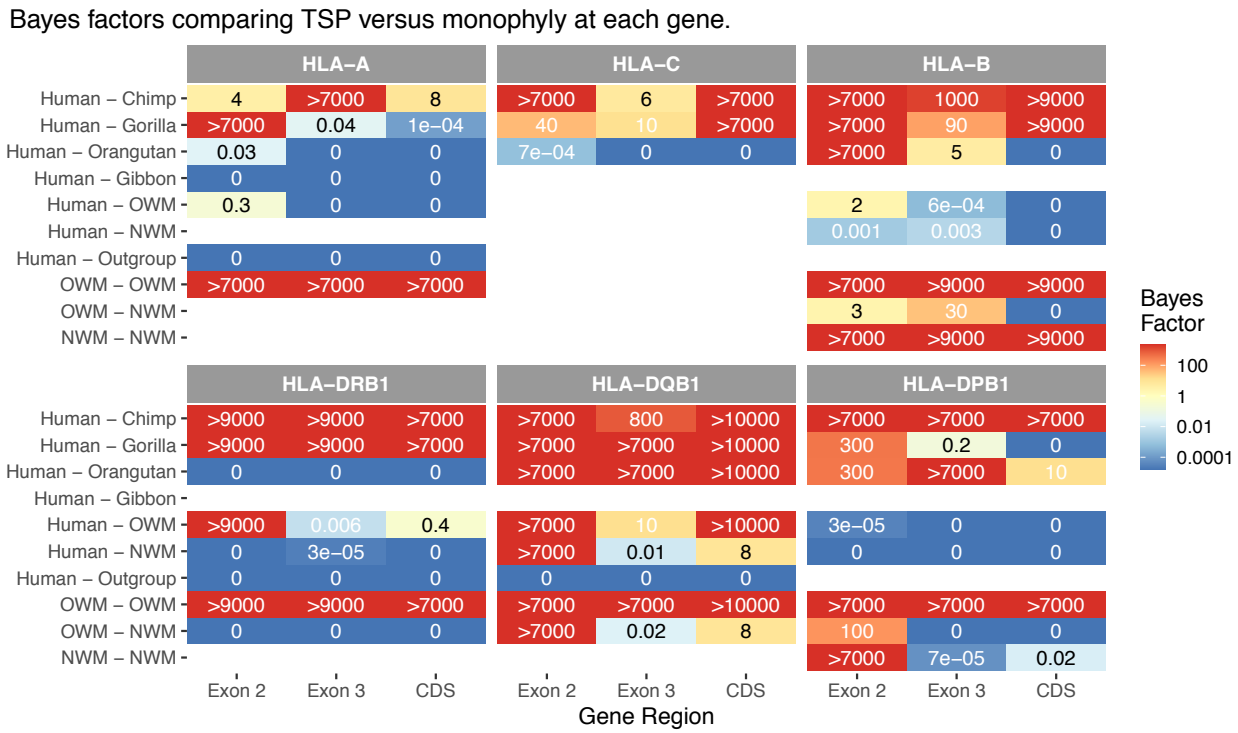

Figure 3: **Strong support for TSP at all six genes.** Bayes factors computed over the set of BEAST trees indicate deep TSP. Different species comparisons are listed on the y-axis, and different gene regions are listed on the x-axis. High Bayes factors (red) indicate support for TSP among the given species for that gene region, while low Bayes factors (blue) indicate that alleles assort according to the species tree, as expected. Bayes factors above 100 are considered decisive. Yellow 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.

tances make less-efficient use of the data compared to the phylogenetic analysis, but should be robust to both convergent nonsynonymous evolution at functional sites and recombination. In the presence of recombination, dS would simply reflect the mean coalescence time along the sequence, averaged over recombined blocks. We estimated dS between pairs of alleles using *codeml*[59,60].

The dS results are shown in Figure 4. Each data point shows the estimated dS for a pair of alleles within or between the indicated species. We reasoned that the observed dS values reflect a distribution of underlying coalescence times between pairs of alleles. Due to the random sampling error in dS, the true distribution of expected values would usually be narrower than the observed distribution. Therefore, for each comparison we estimated the underlying distribution of expected dS values under a simple model (see Supplement for details); these are indicated by the gray boxed region on each plot.

First, we noticed that even within humans, the dS divergence values can be extraordinarily high. Noting that a conservative upper bound on the mutation rate in primates is $\sim 1 \times 10^{-9}$ per year[61], a value of dS > 0.1, as seen in exon 2 for several genes, suggests a remarkably ancient split time of > 50 MY between pairs of human alleles. Second, we reasoned that in the presence of TSP, the distributions of dS values should overlap within and between the relevant species. For example, in the phylogenetic analysis, HLA-A shows evidence for TSP among human, chimpanzee, and gorilla, but not with other species. Consistent with this, we see overlap of the dS distributions for these three species. However, in most cases we actually see *more* evidence for TSP than in the phylogenetic analysis (Figure 4A). For example, HLA-DRB1 shows overlap between the dS distributions of human-human and human-OWM pairs (though they do not quite overlap for chimpanzee, gorilla, and orangutan), while HLA-DQB1 shows broad overlap of the distributions—including the outgroup species.

In Figure 4B, we see overlap of the human-human and human-NWM distributions in exon 2 of HLA-B, indicating deeper TSP than we see in the phylogeny. We do not observe any overlap for the whole CDS, confirming that the greater recombination across larger regions averages dS across blocks and attenuates signal for TSP. Thus, the dS estimates support ancient allelic structure within these genes, consistent with—if not even more ancient than—the results from phylogenetic analysis.

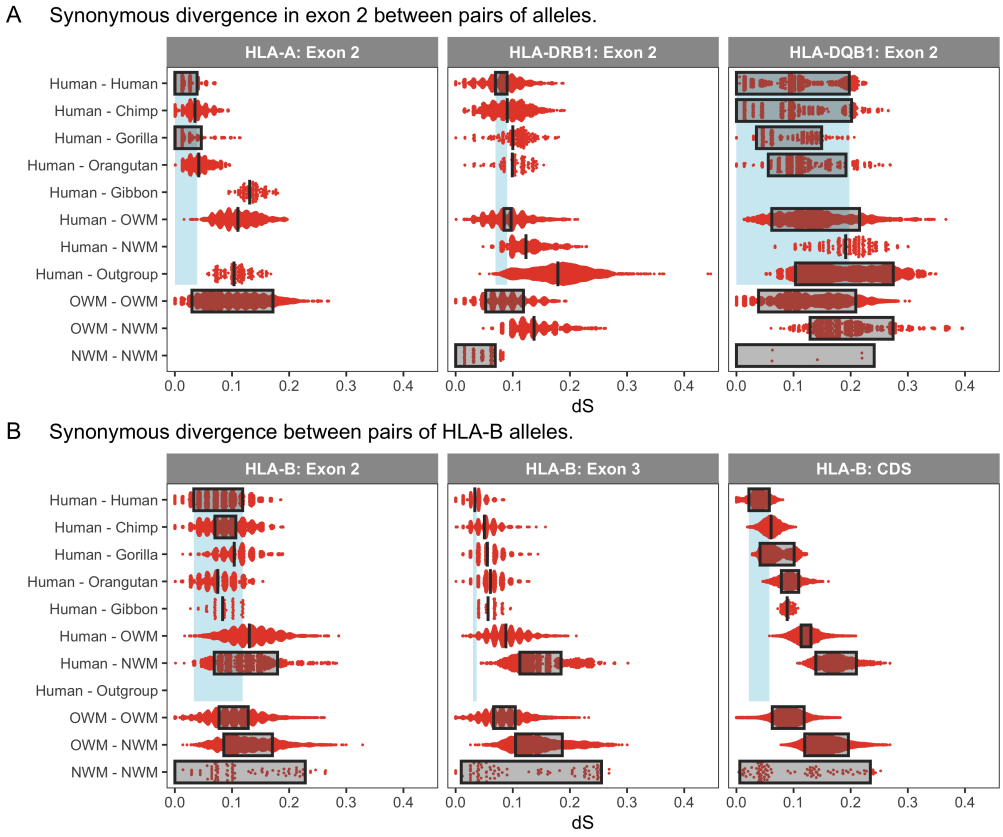

Figure 4: **Synonymous divergence between allele pairs is consistent with ancient TSP.** Each red point represents estimated dS for a pair of alleles (x-axis), sorted by species comparison (y-axis). The gray boxes show estimated ranges of true dS values (details in Supplement) for each comparison. The blue regions extend the gray box for human-human pairs downward across each plot to facilitate visual comparison between species pairs. **A)** In many cases, allele comparisons within humans (Human - Human) show overlapping synonymous divergence values to allele comparisons between humans and other species—for example, between human and OWM alleles in HLA-DRB1. **B)** In HLA-B, the human-human dS distribution overlaps the human-NWM distribution in exon 2, but not in exon 3 or the whole CDS. See Supplementary Figures S16-S20 for dS results for the other genes and regions.

In summary, our data support deep TSP at all six classical MHC genes considered here. The HLA-A, -C, -B, and -DPB1 trees show TSP between human, chimpanzee, and gorilla, while HLA-B and -DPB1 additionally show TSP between human and orangutan. HLA-DRB1 reveals even more extensive TSP, back to our ancestor with OWM. The deepest strongly-supported TSP is at HLA-DQB1, with TSP between humans and NWMs indicating the persistence of ancient lineages for at least 45 million years. The HLA-A, -B, -DRB1, -DQB1, and -DPB1 trees also indicate extensive TSP among OWM. Our dS results corroborate deep TSP at all loci.

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)[9,10] and its large number of GWAS hits for autoimmune and infectious diseases[8,62]. 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. 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 4-fold degenerate sites. The results were insensitive to whether the analysis included the entire CDS or smaller subsets (Supplementary Figure S21).

Figure 5A shows the substitution rate fold-change for each nucleotide along the concatenated coding sequence of HLA-A and -DRB1. Exon boundaries are indicated by red dashed lines. We found that nearly all the rapidly-evolving sites lie within the exons that encode the peptide-binding domains of each protein: for example, 80% of sites evolving at more than 4 times the rate of the 4-fold degenerate sites were located in exons 2 and 3 for the Class I gene HLA-A, and 86% of such sites were located in exon 2 for the Class II gene HLA-DRB1. This pattern holds true for the other Class I and Class II genes as well (Supplementary Figure S22).

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, and mapped these onto the known protein structures. As shown in Figure 5B, rapidly-evolving amino acids (red) 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. Amino acids closer to the peptide have significantly higher evolutionary rates than amino acids further from the peptide, as shown in Figure 5C (see also Supplementary Figures S23-S24). These results are consistent with the expectation that rapid evolution and diversity at the MHC would be mediated by selective pressures for changes in peptide binding.

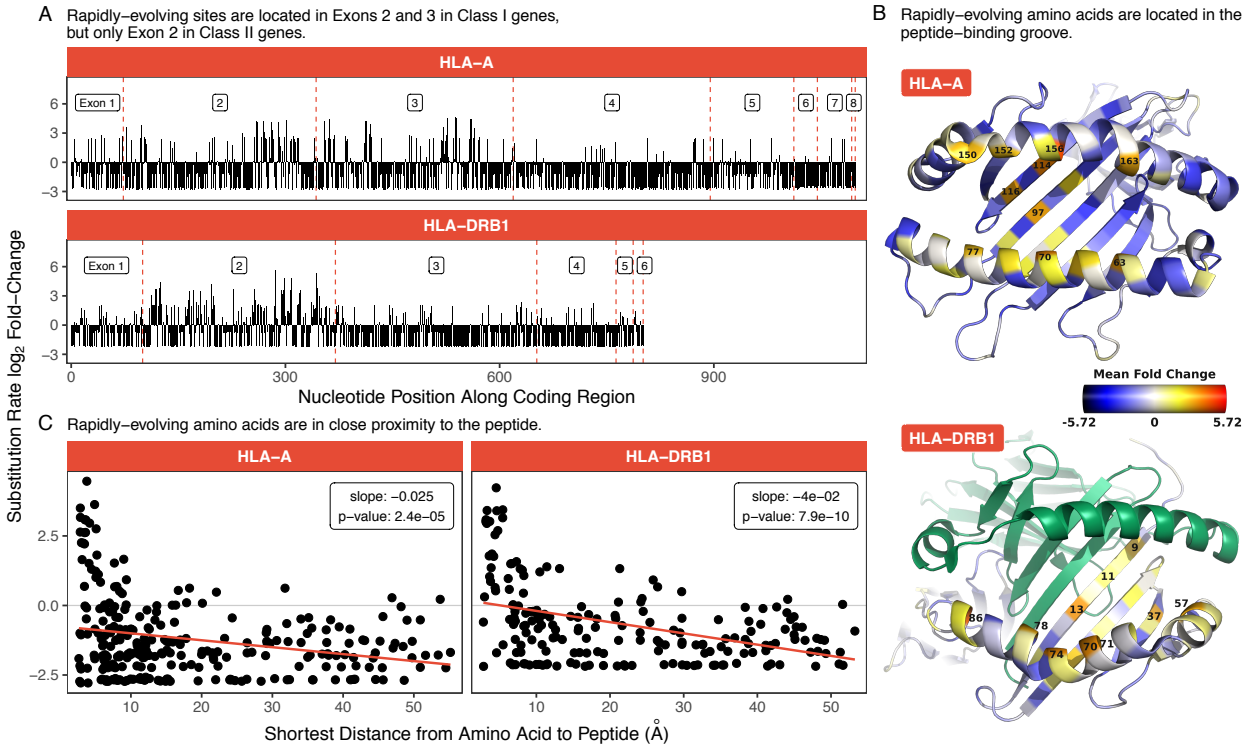

Figure 5: **Rapidly-evolving sites from BEAST2. A)** Rapidly-evolving sites are primarily located in exons 2 and 3. Here, the exons are concatenated such that the cumulative position along the coding region is on the x-axis. The y-axis shows the substitution rate at each site, expressed as a fold-change (the base-2 logarithm of each site's evolutionary rate divided by the average rate among 4-fold degenerate sites within the gene). **B)** Rapidly-evolving sites are located in each protein's peptide-binding pocket. Structures are Protein Data Bank[63] 6D2T[64] for HLA-B and 2IAM[65] for HLA-DRB1, with images created in *PyMOL*[66]. Substitution rates for each amino acid are computed as the mean substitution rate of the three sites composing the codon. Red indicates rapidly-evolving amino acids, while blue indicates conserved amino acids. The top ten fastest-evolving amino acids of each protein are labeled on the structures. **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*[66]. Each point is an amino acid. The red line is a regression of substitution rate on minimum distance, with slope and p-value annotated on each panel.

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 therefore tested whether these sites had known associations with peptide binding, TCR usage, and human disease.

We collected HLA fine-mapping studies for infectious, autoimmune, and other diseases (Supplementary Table S2). 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, as indicated by the colored blocks in Figure 6A. As expected, we found that all six genes harbor many associations with diverse traits and diseases.

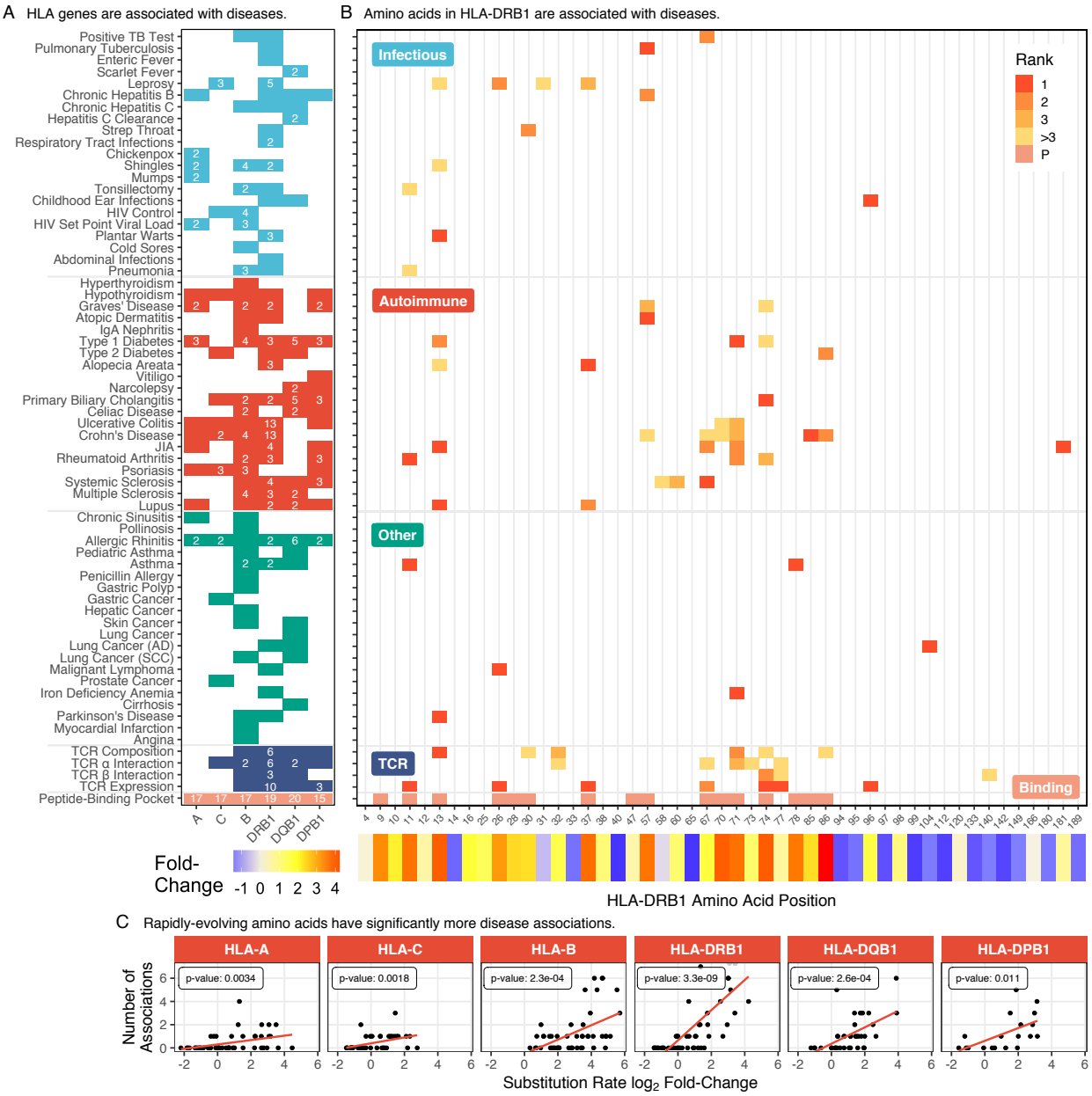

Figure 6: **Disease associations in HLA-DRB1. A)** The six classical HLA genes studied here (x-axis) are associated with many disease and immune-related phenotypes (y-axis). Colored blocks indicate an association between the indicated disease or trait at either SNPs, amino acids, or classical alleles within the corresponding gene. The number of independent associations is indicated when > 1. For the peptide-binding pocket in the bottom row, the printed number indicates the number of critical residues. **B)** Associations between amino acids in HLA-DRB1 and human diseases from HLA fine-mapping studies. The polymorphic amino acid positions ($> 1\%$ MAF) within HLA-DRB1 are on the x-axis, and the diseases and traits are on the y-axis. Associations are colored according to whether they were the top hit, second independent hit, and so on. For peptide binding, a pink box indicates that the amino acid is present (P) among the critical residues. The x-axis is annotated with colored boxes representing *BEAST2* evolutionary rates at codon-level. Red indicates rapidly-evolving positions and blue indicates conserved positions. **C)** The *BEAST2* evolutionary rate is significantly associated with the total number of associations for all six genes.

Moreover, the sites that are rapidly-evolving in the primate analysis show specific enrichment for phenotypic associations in HLA fine-mapping studies. This is illustrated in Figure 6B; the x-axis indicates polymorphic amino acid positions, each annotated with the evolutionary rate from *BEAST2*. The amino acid positions most frequently associated with disease are largely concordant with the rapidly-evolving positions, particularly positions 11, 13, 57, 70, and 71. Supplementary Figures S25-S29 show the associations for the other genes.

Indeed, Figure 5D shows that *BEAST2* evolutionary rate is significantly associated with the number of disease and immune-phenotype associations for all six genes. Together, these results suggest that highly functional sites within the peptide-binding domain are both rapidly-evolving and highly enriched for associations. We wondered whether high variability might be a confounding factor – i.e., that highly polymorphic sites are more likely to harbor significant associations. Indeed, while the number of amino acid alleles at each position is significantly associated with the number of associations, the evolutionary rate remains significant for 5 of the 6 genes in a multiple regression controlling for number of alleles, even though this analysis is highly conservative (Supplementary Table S5). 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 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 nonsynonymous 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.

Despite the 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 hinted at 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 all six studied genes, including between humans and OWM at HLA-DRB1, and even between humans and NWM at HLA-DQB1. Thus, remarkably, lineages at HLA-DQB1 have been maintained for at least 45 million years. The evidence for deep TSP is mainly concentrated within the exons that encode the peptide-binding region for each gene (exons 2 and 3 for Class I, and exon 2 for Class II). Comparisons of synonymous divergence within and between species provide further support for extremely ancient lineages, while obviating concerns about model misspecification in the phylogenetic analysis.

Our evidence for TSP at HLA-DQB1 spanning at least 45 million years places this 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 20 million years[28]. In various chimpanzee species, *OAS1*, which helps inhibit viral replication, contains alleles up to 13my old[67]. TSP between chimpanzee and human includes *LAD1*, a protein that maintains cell cohesion (6my)[68], retroviral transcription factor *TRIM5α* (4-7my in apes and $>$ 8my in OWM)[69,70], and *ZC3HAV1*, an antiviral protein leading to viral RNA degredation (6my)[71], among others[26].

Looking more broadly across the tree of life, ancient trans-species polymorphism occurs widely, albeit rarely. Several of the best examples are found in the MHC locus: MHC polymorphisms have been maintained for 35my in cetaceans[72], 40my in herons[73], 48my in mole rats[74], 70my in tree frogs[75], and over 105my in salmonid fishes[76,77]. 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[78], in

plants, S-genes determining self-incompatibility exhibit TSP spanning 36 million years[79,80], and in *Formica* ants, alleles at a supergene underlying colony queen number have been maintained for over 30 million years[81] (See the Supplement for more examples).

Paradoxically, given the extremely long-lived balancing selection acting in these lineages, many authors have also reported strong directional selection[82–84]. Indeed, within the phylogeny we find that the most rapidly-evolving codons are substituted at around 5–7-fold the neutral rate. For all six genes, these rapidly-evolving sites lie within the peptide binding regions of the corresponding proteins, usually very close to the peptide-contact surfaces. Moreover, the primary role of 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.

We further connected our evolutionary results to their functional consequences by examining published associations between HLA variants and immune-related phenotypes. The peptide-presentation and recognition process is integral to the immune response and affects our ability to fight infections and recognize self-antigens. After collating HLA fine-mapping studies of human disease, including infectious, autoimmune, and other diseases, we found that the same rapidly-evolving amino acids are also key sites of disease association.

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. Because it is challenging to detect pathogens with both specificity and sensitivity, many MHC alleles also happen to bind self-peptides. This results in rapidly-evolving amino acids being associated with both infections and autoimmune conditions, as well as TCR phenotypes and peptide-binding residues.

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 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[1,85].

Materials and Methods

Data. MHC allele sequences for HLA-A, -B, -C, -DPB1, -DQB1, and -DRB1 were downloaded from the IPD Database[6,7]. The database contains tens of thousands of alleles, but we selected only the alleles for which the entire coding sequence (all exons) was available. In humans, this consisted of 2,819 HLA-A alleles, 3,338 HLA-B alleles, 3,055 HLA-C alleles, 522 HLA-DPB1 alleles, 297 HLA-DQB1 alleles, and 200 HLA-DRB1 alleles. Excluding humans, there were 14 species and 240 alleles for HLA-A, 24 species and 1,334 alleles for HLA-B, 6 species and 61 alleles for HLA-C, 13 species and 185 alleles for HLA-DPB1, 16 species and 288 alleles for HLA-DQB1, and 15 species and 231 alleles for HLA-DRB1. Note that we use the familiar human-centric names (with prefix "HLA-") to refer to these genes in all species, although the gene names with prefix "MHC-" are more accurate in an evolutionary context.

BEAST2 is computationally limited by the number of sequences. Thus, when hundreds or thousands of alleles were available for a single species, we reduced each set to a representative set spanning the species' allelic diversity, according to the alleles' two-digit designations. The final set consisted of 149 alleles for HLA-A, 197 for HLA-B, 61 for HLA-C, 100 for HLA-DPB1, 122 for HLA-DQB1, and 122 for HLA-DRB1 (lists of alleles provided as Supplementary Files).

Alleles for each gene were aligned using *MUSCLE*[86] in *MEGA X*[87] with default settings.

Modern human data was obtained for 2504 unrelated individuals from Phase 3 of the 1000 Genomes Project, from the re-sequencing done at the New York Genome Center and mapped to GRCh38[88,89].

Nucleotide Diversity. The classical MHC region is defined as chr6:28,510,120-33,480,577 (GRCh38)[2]. Nucleotide diversity (π) was calculated on the modern human data using *VCFtools (0.1.15)*[90]. For the entire MHC region (Figure 1A), π was calculated in 5000bp sliding windows with a step size of 1000bp. For each gene separately (Figure 1B-C and Supplementary Figure S1), π was calculated in 50bp sliding windows with a step size of 10bp.

Bayesian Phylogenetic Analysis. We constructed phylogenetic trees using *BEAST2*[53,54] with package *SubstBMA*[55]. *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.[55]. 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. Even though the substitution rate is known to be different over the evolutionary time we consider, Wu et al.[55] demonstrated virtually no difference in likelihoods or site partitions when using a strict vs. relaxed molecular clock. Thus, we used a strict clock to further simplify our analyses.

Priors. For the Dirichlet process priors, we used the informative priors constructed by Wu et al.[55] 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 ran *BEAST2* on allele sequences from the IPD Database[5–7], considering subsets of the coding sequence: 1) the entire coding sequence, 2) exon 2 only, 3) exon 3 only,

4) exons 2 and 3 together (the peptide-binding region), and 5) all other exons (excluding exons 2 and 3). Because *BEAST2* runtime scales with the number of samples, we restricted the number of alleles used for each gene. We only considered alleles that 1) have the entire coding sequence available in the database (because many submitted alleles are the result of sequencing exon 2 and 3 alone), 2) do not have a nonsense mutation, and 3) adequately represent the diversity of the species. For example, we restricted human alleles to include at least one allele representing every major type (2-field name). For macaques, which have even more MHC diversity than humans, we limited alleles to 5 randomly-chosen alleles per species that also span several types.

The xml files we used to run *BEAST2* were based closely on those used for the mammal dataset with RDPM model and strict clock in Wu et al.[55] (`https://github.com/jessiewu/substBMA/blob/master/examples/mammal/mammal_rdpm_sc.xml`). However, due to the complexity of the MHC, we ran each for 40,000,000 states (instead of 25,000,000), using the first 10% as burn-in and sampling every 10,000 states. The xml files required to run all our analyses are provided as Supplementary Files. Additionally, we ran 4-12 replicates (4 at a time) for each of the 6 genes and each of the 5 subsets of the coding sequence, until we obtained at least 4 replicates whose likelihood and posterior parameter distributions all agreed, as recommended by *BEAST2* and explored in *Tracer* version 1.7.1[91]. This ensures that all of the replicates were exploring the same parameter space and were converging upon the same global optimum, allowing the $\geq$ 4 independent runs to be justifiably combined. We combined the matching replicates using *LogCombiner* version 2.6.3[92], which aggregates the results across all states. We then used the combined results to perform our analyses.

HLA-B. Running *BEAST2* on HLA-B proved difficult due to the number of sequences included as well as the considerable diversity among them. To improve convergence, we employed coupled MCMC in *BEAST2* for the entire coding sequence (subset 1) and exons 2 and 3 combined (subset 4). 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[93]. Using coupled MCMC sped up these more-difficult *BEAST2* runs and allowed us to obtain 4 valid replicates that could be used for the analysis. Xml files for these runs are also provided as Supplementary Files.

Phylogenetic Trees. Since we sampled every 10,000 states for each *BEAST2* replicate, discarded the first 10% as burn-in, and obtained 4-7 acceptable replicates per gene/sequence subset, we obtained 14,101 - 25,207 phylogenies per gene/sequence subset. We used *TreeAnnotator* version 2.6.3[92] to summarize each set of possible trees as a maximum clade credibility tree, which is the tree that maximizes the product 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[56]. 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.)

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[58].

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 comparisons, including sets chosen at random and sets that appear to support TSP in the *BEAST2* summary trees. 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 group and 2 possible unrooted trees where the 2 human alleles would not form a monophyletic group, as shown in Figure 7.

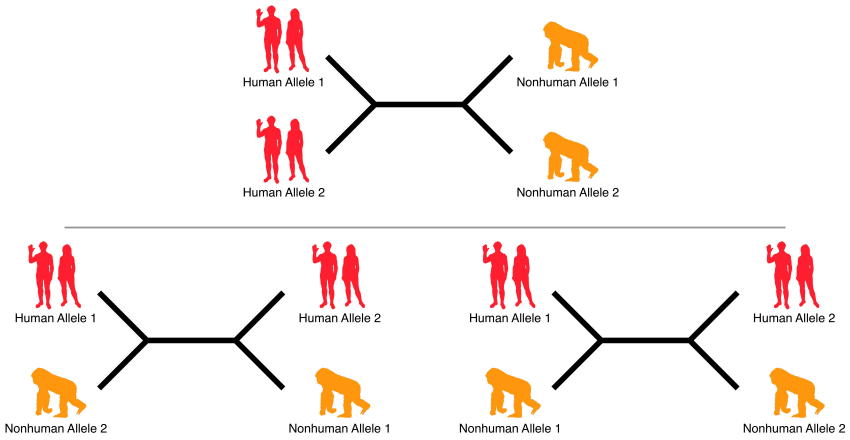

Figure 7: **Possible unrooted trees of 4 alleles.** There is one tree where the human alleles are monophyletic, and two trees where they are non-monophyletic.

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 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[58].

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

For each gene and genic region, we tested for TSP between human and chimpanzee, gorilla, orangutan, gibbon, OWM, and NWM. We also computed TSP for groups not involving humans, including among OWM themselves, between OWM and NWM, and among NWM themselves. Lastly, we tested for TSP between humans and the outgroups, for which we did not expect to see any evidence of TSP.

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. To normalize the rates, we divided them by the average evolutionary rate among 4-fold degenerate sites. 4-fold degenerate sites for each gene were identified from our *codeml* analysis using the entire coding sequence.

We expressed the normalized per-site rates as fold-changes by taking the base-2 logarithm. We calculated per-amino-acid rates by averaging the per-site rates among the three sites composing each codon.

Protein Structures. We used *PyMOL* version 2.4.2[66] to visualize the per-codon evolutionary rates on each gene's protein structure. We used model 4F7P[94] from Protein Data Bank[63] (`https://www.rcsb.org/`) for HLA-A, 6D2T[64] for HLA-B, 4NT6[95] for HLA-C, 4P4K[96] for HLA-DPB1, 1UVQ[97] for HLA-DQB1, and 2IAM[65] for HLA-DRB1.

We calculated the distances between all amino acids of the HLA molecule and all amino acids of the peptide in *PyMOL*, then took the minimum distance to represent that amino acid's overall distance to the peptide. We averaged the minimum distances over three alternative structures for each gene, to prevent relying too heavily on a particular structure. We used models 4F7P[94], 6J1V[98], and 3MGO[99] from Protein Data Bank[63] for HLA-A; 6D2T[64], 3BVN[100], and 6PYJ[101] for HLA-B; 4NT6[95], 5W67[102], and 5VGE[103] for HLA-C; 4P4K[96], 3LQZ[104], and 3WEX[105] for HLA-DPB1; 1UVQ[97], 2NNA[106], and 4D8P[107] for HLA-DQB1; and 2IAM[65], 6ATF[108], and 5LAX[109] for HLA-DRB1.

Synonymous Diversity. Synonymous diversity of the IPD alleles was calculated pairwise using *codeml*[59,60] on subsets of the coding sequence. We used a larger set of alleles than we used in *BEAST2*, because codeml is not as computationally limited by the number of samples. We used all alleles that both 1) have the entire coding sequence available in the database (because many submitted alleles are the result of sequencing exons 2 and 3 alone) and 2) do not have a nonsense mutation. This set consisted of 289 alleles for HLA-A, 423 for HLA-B, 138 for HLA-C, 255 for HLA-DPB1, 352 for HLA-DQB1, and 287 for HLA-DRB1 (lists of alleles provided as Supplementary Files). This larger set is a superset of the alleles used for the *BEAST2* analysis. We ran *codeml* using runmode=−2 for pairwise calculation (to avoid relying on a species tree), seqtype=1 for codon-based, CodonFreq=1 to estimate the equilibrium codon frequencies from the average nucleotide frequencies, model=0 and NSsites=0 to specify the basic codon substitution model, fix_kappa=0 to estimate the κ parameter with initial kappa=3 for all genes except HLA-C, where initial kappa=2, and fix_omega=0 to estimate the ω parameter with initial omega=0.5. *codeml* also outputted 4-fold degenerate sites, which we used for normalization of the *BEAST2* rates, as described above.

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 1000 cases and which identified putatively independent signals via conditional analysis. We included all independent signals identified as significant by the original authors, regardless of whether they were entire alleles, amino acids, or SNPs. If there was more than one study for the same disease, but in different populations, we included all unique independent hits. References for studies are listed in Supplementary Table S2. We also collected associations between amino acids and TCR phenotypes, with references listed in Supplementary Table S3. References for the anchor residues that make up the peptide-binding pocket are listed in Supplementary Table S4.

Polymorphic Amino Acid Positions. Amino acid allele frequencies were obtained from Luo et al.[110], who imputed HLA variation in 1000G using *SNP2HLA* (`https://github.com/immunogenomics/HLA-TAPAS/blob/master/resources/1000G.bglv4.FRQ.frq`). Amino acid positions with MAF> 1% were considered polymorphic for the purposes of plotting.

Acknowledgements

We thank Karla Kirkegaard, Dmitri Petrov, Molly Przeworski, Soumya Raychaudhuri, Noah Rosenberg, Molly Schumer, and members of the Pritchard lab for helpful discussions. This work was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1656518 for ALF and by R01HG008140.

References

[1] Jacek Radwan, Wiesław Babik, Jim Kaufman, Tobias L. Lenz, and Jamie Winternitz. Advances in the Evolutionary Understanding of MHC Polymorphism. *Trends in Genetics*, 36(4):298–311, 2020. ISSN 13624555. doi: 10.1016/j.tig.2020.01.008.

[2] Human Genome Region MHC, 2022. URL `https://www.ncbi.nlm.nih.gov/grc/human/regions/MHC`.

[3] Nuala A. O'Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O'Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, and Kim D. Pruitt. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745, 2016. ISSN 13624962. doi: 10.1093/nar/gkv1189.

[4] Jacques Neefjes, Marlieke L.M. Jongsma, Petra Paul, and Oddmund Bakke. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11 (12):823–836, 2011. ISSN 14741733. doi: 10.1038/nri3084.

[5] Giuseppe Maccari, James Robinson, Keith Ballingall, Lisbeth A. Guethlein, Unni Grimholt, Jim Kaufman, Chak Sum Ho, Natasja G. De Groot, Paul Flicek, Ronald E. Bontrop, John A. Hammond, and Steven G.E. Marsh. IPD-MHC 2.0: An improved inter-species database for the study of the major histocompatibility complex. *Nucleic Acids Research*, 45(D1):D860–D864, 2017. ISSN 13624962. doi: 10.1093/nar/gkw1050.

[6] Giuseppe Maccari, James Robinson, John A. Hammond, and Steven G.E. Marsh. The IPD Project: a centralised resource for the study of polymorphism in genes of the immune system. *Immunogenetics*, 72(1):49–55, 2020. ISSN 14321211. doi: 10.1007/s00251-019-01133-w.

[7] James Robinson, Dominic J. Barker, Xenia Georgiou, Michael A. Cooper, Paul Flicek, and Steven G.E. Marsh. IPD-IMGT/HLA Database. *Nucleic Acids Research*, 48(D1):D948–D955, 2020. ISSN 13624962. doi: 10.1093/nar/gkz950.

[8] A Buniello, JAL MacArthur, M Cerezo, LW Harris, J Hayhurst, C Malangone, A McMahon, J Morales, E Mountjoy, E Sollis, D Suveges, O Vrousgou, PL Whetzel, R Amode, JA Guillen, HS Riat, SJ Trevanion, P Hall, H Junkins, P Flicek, T Burdett, LA Hindorff, F Cunningham, and H Parkinson. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Research*, 47(D1):D1005–D1012, 2019. doi: 10.1093/nar/gky1120.

[9] Austin L. Hughes and Masatoshi Nei. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveal overdominant selection. *Nature*, 335(6186):167–170, 1988. doi: 10.1038/335167a0.

[10] A. L. Hughes and M. Nei. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences of the United States of America*, 86(3):958–962, 1989. ISSN 00278424. doi: 10.1073/pnas.86.3.958.

[11] B. Arden and J. Klein. 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*, 79(7):2342–2346, 1982. ISSN 00278424. doi: 10.1073/pnas.79.7.2342.

[12] W. E. Mayer, M. Jonker, D. Klein, P. Ivanyi, G. van Seventer, and J. Klein. Nucleotide sequences of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *The EMBO journal*, 7 (9):2765–2774, 1988. ISSN 02614189. doi: 10.1002/j.1460-2075.1988.tb03131.x.

[13] Edward K. Wakeland, Roy W Tarruzzee, Cheng-Chan Lu, Wayne Potts, Richard A McIndoe, William S Talbot, and Thomas J McConnell. The evolution of MHC class II genes within the genus Mus. In Chella S. David, editor, *H-2 Antigens: Genes, Molecules, Function*, pages 139–153. Springer, Boston, MA, 1 edition, 1987. ISBN 978-1-4757-0766-3. doi: 10.1007/978-1-4757-0764-9{_}14.

[14] Masatoshi Nei and Austin L Hughes. Polymorphism and evolution of the major histocompatibility complex loci in mammals. In R Selander, A Clark, and T Whittam, editors, *Evolution at the Molecular Level*, chapter 11, pages 222–247. Sinauer Associates, Inc, Sunderland, MA, 6 edition, 1991. ISBN 978-0878938193. doi: 10.1093/oxfordjournals.molbev.a040698.

[15] Iain Mathieson, Iosif Lazaridis, Nadin Rohland, Swapan Mallick, Nick Patterson, Songül Alpaslan Roodenberg, Eadaoin Harney, Kristin Stewardson, Daniel Fernandes, Mario Novak, Kendra Sirak, Cristina Gamba, Eppie R. Jones, Bastien Llamas, Stanislav Dryomov, Joseph Pickrell, Juan Luís Arsuaga, José María Bermúdez De Castro, Eudald Carbonell, Fokke Gerritsen, Aleksandr Khokhlov, Pavel Kuznetsov, Marina Lozano, Harald Meller, Oleg Mochalov, Vyacheslav Moiseyev, Manuel A.Rojo Guerra, Jacob Roodenberg, Josep Maria Vergès, Johannes Krause, Alan Cooper, Kurt W. Alt, Dorcas Brown, David Anthony, Carles Lalueza-Fox, Wolfgang Haak, Ron Pinhasi, and David Reich. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528(7583):499–503, 2015. ISSN 14764687. doi: 10.1038/nature16152.

[16] Yair Field, Evan A. Boyle, Natalie Telis, Ziyue Gao, Kyle J. Gaulton, David Golan, Loic Yengo, Ghislain Rocheleau, Philippe Froguel, Mark I. McCarthy, and Jonathan K. Pritchard. Detection of human adaptation during the past 2000 years. *Science*, 354(6313):760–764, 2016. ISSN 10959203. doi: 10.1126/science.aag0776.

[17] Morten E Allentoft, Martin Sikora, Alba Refoyo-martínez, Evan K Irving-pease, and Anders Fischer. Population Genomics of Stone Age Eurasia. *bioRxiv*, 2022. doi: 10.1101/2022.05.04.490594.

[18] Pei-kuan Cong, Wei-Yang Bai, Jin-Chen Li, Meng-Yuan Yang, Saber Khederzadeh, Si-Rui Gai, Nan Li, Yu-Heng Liu, Shi-Hui Yu, Wei-Wei Zhao, Jun-Quan Liu, Yi Sun, Xiao-Wei Zhu, Pian-Pian Zhao, Jiang-Wei Xia, Peng-Lin Guan, Yu Qian, Jian-Guo Tao, Lin Xu, Geng Tian, Ping-Yu Wang, Shu-Yang Xie, Mo-Chang Qiu, Ke-Qi Liu, Bei-Sha Tang, and Hou-Feng Zheng. Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nature Communications*, 13(1):1–15, 2022. doi: 10.1038/s41467-022-30526-x.

[19] Yukinori Okada, Yukihide Momozawa, Saori Sakaue, Masahiro Kanai, Kazuyoshi Ishigaki, Masato Akiyama, Toshihiro Kishikawa, Yasumichi Arai, Takashi Sasaki, Kenjiro Kosaki, Makoto Suematsu, Koichi Matsuda, Kazuhiko Yamamoto, Michiaki Kubo, Nobuyoshi Hirose, and Yoichiro Kamatani. Deep whole-genome sequencing reveals recent selection signatures linked to evolution and disease risk of Japanese. *Nature Communications*, 9(1):1–10, 2018. ISSN 20411723. doi: 10.1038/s41467-018-03274-0.

[20] Yoshiaki Yasumizu, Saori Sakaue, Takahiro Konuma, Ken Suzuki, Koichi Matsuda, Yoshinori Murakami, Michiaki Kubo, Pier Francesco Palamara, Yoichiro Kamatani, Yukinori Okada, and Yoko Satta. Genome-Wide Natural Selection Signatures Are Linked to Genetic Risk of Modern Phenotypes in the Japanese Population. *Molecular Biology and Evolution*, 37(5):1306–1316, 2020. ISSN 15371719. doi: 10.1093/molbev/msaa005.

[21] Ravi Sachidanandam, David Weissman, Steven C. Schmidt, Jerzy M. Kakol, Lincoln D. Stein, Gabor Marth, Steve Sherry, James C. Mullikin, Beverley J. Mortimore, David L. Willey, Sarah E. Hunt, Charlotte G. Cole, Penny C. Coggill, Catherine M. Rice, Zemin Ning, Jane Rogers, David R. Bentley, Pui Yan Kwok, Elaine R. Mardis, Raymond T. Yeh, Brian Schultz, Lisa Cook, Ruth Davenport, Michael Dante, Lucinda Fulton, Ladeana Hillier, Robert H. Waterston, John D. McPherson,

Brian Gilman, Stephen Schaffner, William J. Van Etten, David Reich, John Higgins, Mark J. Daly, Brendan Blumenstiel, Jennifer Baldwin, Nicole Stange-Thomann, Michael C. Zody, Lauren Linton, Eric S. Lander, and David Altshuler. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822):928–933, 2001. ISSN 00280836. doi: 10.1038/35057149.

[22] Dieter Ebert and Peter D. Fields. Host–parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, 21(12):754–768, 2020. ISSN 14710064. doi: 10.1038/s41576-020-0269-1.

[23] Federica Pierini and Tobias L. Lenz. Divergent allele advantage at human MHC genes: Signatures of past and ongoing selection. *Molecular Biology and Evolution*, 35(9):2145–2158, 2018. ISSN 15371719. doi: 10.1093/molbev/msy116.

[24] N. Takahata and M. Nei. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*, 124(4):967–978, 1990. ISSN 00166731. doi: 10.1093/genetics/124.4.967.

[25] Swapan Mallick, Heng Li, Mark Lipson, Iain Mathieson, Melissa Gymrek, Fernando Racimo, Mengyao Zhao, Niru Chennagiri, Susanne Nordenfelt, Arti Tandon, Pontus Skoglund, Iosif Lazaridis, Sriram Sankararaman, Qiaomei Fu, Nadin Rohland, Gabriel Renaud, Yaniv Erlich, Thomas Willems, Carla Gallo, Jeffrey P. Spence, Yun S. Song, Giovanni Poletti, Francois Balloux, George Van Driem, Peter De Knijff, Irene Gallego Romero, Aashish R. Jha, Doron M. Behar, Claudio M. Bravi, Cristian Capelli, Tor Hervig, Andres Moreno-Estrada, Olga L. Posukh, Elena Balanovska, Oleg Balanovsky, Sena Karachanak-Yankova, Hovhannes Sahakyan, Draga Toncheva, Levon Yepiskoposyan, Chris Tyler-Smith, Yali Xue, M. Syafiq Abdullah, Andres Ruiz-Linares, Cynthia M. Beall, Anna Di Rienzo, Choongwon Jeong, Elena B. Starikovskaya, Ene Metspalu, Jüri Parik, Richard Villems, Brenna M. Henn, Ugur Hodoglugil, Robert Mahley, Antti Sajantila, George Stamatoyannopoulos, Joseph T.S. Wee, Rita Khusainova, Elza Khusnutdinova, Sergey Litvinov, George Ayodo, David Comas, Michael F. Hammer, Toomas Kivisild, William Klitz, Cheryl A. Winkler, Damian Labuda, Michael Bamshad, Lynn B. Jorde, Sarah A. Tishkoff, W. Scott Watkins, Mait Metspalu, Stanislav Dryomov, Rem Sukernik, Lalji Singh, Kumarasamy Thangaraj, Svante Paäbo, Janet Kelso, Nick Patterson, and David Reich. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature*, 538(7624):201–206, 2016. ISSN 14764687. doi: 10.1038/nature18964.

[26] Ellen M. Leffler, Ziyue Gao, Susanne Pfeifee, Laure Ségurel, Adam Auton, Oliver Venn, Rory Bowden, Ronald Bontrop, Jeffrey D. Wall, Guy Sella, Peter Donnelly, Gilean McVean, and Molly Przeworski. Multiple Instances of Ancient Balancing Selection Shared Between Humans and Chimpanzees. *Science*, 339(6127):1578–1582, 2013. doi: 10.1126/science.1234070.

[27] Luisa Azevedo, Catarina Serrano, Antonio Amorim, and David N. Cooper. Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. *Human Genomics*, 9(1):4–9, 2015. ISSN 14797364. doi: 10.1186/s40246-015-0043-1.

[28] Laure Ségurel, Emma E. Thompson, Timothée Flutre, Jessica Lovstad, Aarti Venkat, Susan W. Margulis, Jill Moyse, Steve Ross, Kathryn Gamble, Guy Sella, Carole Ober, and Molly Przeworski. The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45):18493–18498, 2012. ISSN 10916490. doi: 10.1073/pnas.1210603109.

[29] Jan Klein. Generation of diversity at MHC loci: Implications for T-cell receptor repertoires. In M. Fougereau and J. Dausset, editors, *Immunology*, volume 80, pages 239–253. Academic Press, London, 1980.

[30] Jan Klein. Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Human Immunology*, 19(3):155–162, 1987. ISSN 01988859. doi: 10.1016/0198-8859(87)90066-8.

[31] Felipe Figueroa, Eberhardt Gúnther, and Jan Klein. MHC polymorphism pre-dating speciation. *Nature*, 335(6187):265–267, 1988. ISSN 00280836. doi: 10.1038/335265a0.

[32] Thomas J McConnell, William S Talbot, Richard A McIndoe, and Edward K Wakeland. The origin of MHC class II gene polymorphism in the genus Mus. *Nature*, 332(14):651–654, 1988. ISSN 10958312. doi: 10.1111/j.1095-8312.1990.tb00826.x.

[33] David A. Lawlor, Frances E. Ward, Peter D. Ennis, Antony P. Jackson, and Peter Parham. HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. *Nature*, 335(6187):268–271, 1988. ISSN 00280836. doi: 10.1038/335268a0.

[34] Bastiaan L. Slierendregt, Nel Otting, Marcel Kenter, and Ronald E. Bontrop. Allelic diversity at the Mhc-DP locus in rhesus macaques (Macaca mulatta). *Immunogenetics*, 41(1):29–37, 1995. ISSN 00937711. doi: 10.1007/BF00188429.

[35] Jonathan E. Boyson, Clare Shufflebotham, Luis F. Cadavid, Julie A. Urvater, Leslie A. Knapp, Austin L. Hughes, and David I. Watkins. 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*, 156(12): 4656–4665, 1996. ISSN 1550-6606.

[36] L. M. McKenzie, J. Pecon-Slattery, M. Carrington, and S. J. O'Brien. Taxonomic hierarchy of HLA class I allele sequences. *Nature Genes and Immunity*, 1(2):120–129, 1999. ISSN 14664879. doi: 10.1038/sj.gene.6363648.

[37] Emily E. Wroblewski, Lisbeth A. Guethlein, Paul J. Norman, Yingying Li, Christiana M. Shaw, Alex S. Han, Jean-Bosco N. Ndjango, Steve Ahuka-Mundeke, Alexander V. Georgiev, Martine Peeters, Beatrice H. Hahn, and Peter Parham. Bonobos Maintain Immune System Diversity with Three Functional Types of MHC-B. *The Journal of Immunology*, 198(9):3480–3493, 2017. ISSN 0022-1767. doi: 10.4049/jimmunol.1601955.

[38] Werner E. Mayer, Colm O'hUigin, Zofia Zaleska-Rutczynska, and Jan Klein. Trans-species origin of Mhc-DRB polymorphism in the chimpanzee. *Immunogenetics*, 37(1):12–23, 1992. ISSN 14321211. doi: 10.1007/BF00223540.

[39] Uwe Brändle, Hideki Ono, Vladimir Vincek, Dagmar Klein, Mladen Golubic, Blazenka Grahovac, and Jan Klein. Trans-species evolution of Mhc-DRB haplotype polymorphism in primates: Organization of DRB genes in the chimpanzee. *Immunogenetics*, 36(1):39–48, 1992. ISSN 14321211. doi: 10.1007/BF00209291.

[40] Heike Kupfermann, Werner E. Mayer, Colm O'hUigin, Dagmar Klein, and Jan Klein. Shared polymorphism between gorilla and human major histocompatibility complex DRB loci. *Human Immunology*, 34(4):267–278, 1992. ISSN 01988859. doi: 10.1016/0198-8859(92)90026-J.

[41] Bastiaan L. Slierendregt, Jacqueline T. van Noort, Ruud M. Bakas, Nel Otting, Margreet Jonker, and Ronald E. Bontrop. Evolutionary stability of transspecies major histocompatibility complex class II DRB lineages in humans and rhesus monkeys. *Human Immunology*, 35(1):29–39, 1992. ISSN 01988859. doi: 10.1016/0198-8859(92)90092-2.

[42] Annemieke Geluk, Diënne G. Elferink, Bastiaan L. Slierendregt, Krista E. Van Meijgaarden, René R.P. Vries, Tom H.M. Ottenhoff, and Ronald E. Bontrop. Evolutionary conservation of major histocompatibility complex-DR/peptide/T cell interactions in primates. *Journal of Experimental Medicine*, 177(4):979–987, 1993. ISSN 15409538. doi: 10.1084/jem.177.4.979.

[43] Yoko Satta, Werner E. Mayer, and Jan Klein. Evolutionary relationship of HLA-DRB genes inferred from intron sequences. *Journal of Molecular Evolution*, 42(6):648–657, 1996. ISSN 00222844. doi: 10.1007/BF02338798.

[44] N. Otting, N. G. De Groot, M. C. Noort, G. G.M. Doxiadis, and R. E. Bontrop. Allelic diversity of Mhc-DRB alleles in rhesus macaques. *Tissue Antigens*, 56(1):58–68, 2000. ISSN 00012815. doi: 10.1034/j.1399-0039.2000.560108.x.

[45] Karin Kriener, Colm O'hUigin, Herbert Tichy, and Jan Klein. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics*, 51(3): 169–178, 2000. ISSN 00937711. doi: 10.1007/s002510050028.

[46] U. B. Gyllensten, D. Lashkari, and H. A. Erlich. Allelic diversification at the class II DQB locus of the mammalian major histocompatibility complex. *Proceedings of the National Academy of Sciences of the United States of America*, 87(5):1835–1839, 1990. ISSN 00278424. doi: 10.1073/pnas.87.5.1835.

[47] Karin Kriener, Colm O'hUigin, and Jan Klein. Independent origin of functional MHC class II genes in humans and New World monkeys. *Human Immunology*, 62(1):1–14, 2001. ISSN 01988859. doi: 10.1016/S0198-8859(00)00233-0.

[48] N. Otting, M. Kenter, P. Van Weeren, M. Jonker, and R. E. Bontrop. Mhc-DQB repertoire variation in hominoid and old world primate species. *Journal of Immunology*, 149(2):461–470, 1992. ISSN 0022-1767.

[49] Nel Otting and Ronald E. Bontrop. Evolution of the major histocompatibility complex DPA1 locus in primates. *Human Immunology*, 42(2):184–187, 1995. ISSN 01988859. doi: 10.1016/0198-8859(94)00095-8.

[50] Nel Otting, Natasja G. De Groot, Gaby G.M. Doxiadis, and Ronald E. Bontrop. Extensive Mhc-DQB variation in humans and non-human primate species. *Immunogenetics*, 54(4):230–239, 2002. ISSN 00937711. doi: 10.1007/s00251-002-0461-9.

[51] IPD-IMGT/HLA Statistics, 2022. URL `https://www.ebi.ac.uk/ipd/imgt/hla/about/statistics/`.

[52] IPD-MHC Statistics, 2022. URL `https://www.ebi.ac.uk/ipd/mhc/statistics/`.

[53] Remco Bouckaert, Joseph Heled, Denise Kühnert, Tim Vaughan, Chieh Hsi Wu, Dong Xie, Marc A. Suchard, Andrew Rambaut, and Alexei J. Drummond. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Computational Biology*, 10(4):1–6, 2014. ISSN 15537358. doi: 10.1371/journal.pcbi.1003537.

[54] Remco Bouckaert, Timothy G. Vaughan, Joëlle Barido-Sottani, Sebastián Duchêne, Mathieu Fourment, Alexandra Gavryushkina, Joseph Heled, Graham Jones, Denise Kühnert, Nicola De Maio, Michael Matschiner, Fábio K. Mendes, Nicola F. Müller, Huw A. Ogilvie, Louis Du Plessis, Alex Popinga, Andrew Rambaut, David Rasmussen, Igor Siveroni, Marc A. Suchard, Chieh Hsi Wu, Dong Xie, Chi Zhang, Tanja Stadler, and Alexei J. Drummond. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4):1–28, 2019. ISSN 15537358. doi: 10.1371/journal.pcbi.1006650.

[55] Chieh Hsi Wu, Marc A. Suchard, and Alexei J. Drummond. Bayesian selection of nucleotide substitution models and their site assignments. *Molecular Biology and Evolution*, 30(3):669–688, 2013. ISSN 07374038. doi: 10.1093/molbev/mss258.

[56] SUMMARIZING POSTERIOR TREES. URL `https://www.beast2.org/summarizing-posterior-trees/`.

[57] HLA Nomenclature, 2019. URL `http://hla.alleles.org/nomenclature/naming.html`.

[58] Harold Jeffreys. *The Theory of Probability*. OUP Oxford, 3 edition, 1998. ISBN 9780191589676.

[59] Ziheng Yang. Paml: A program package for phylogenetic analysis by maximum likelihood. *Computer Applications in BioSciences*, 13(5):555–556, 1997. ISSN 14602059. doi: 10.1093/bioinformatics/13.5.555.

[60] Ziheng Yang. PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, 24(8):1586–1591, 2007. ISSN 07374038. doi: 10.1093/molbev/msm088.

[61] Manjusha Chintalapati and Priya Moorjani. Evolution of the mutation rate across primates. *Current Opinion in Genetics and Development*, 62:58–64, 2020. ISSN 18790380. doi: 10.1016/j.gde.2020.05.028.

[62] A. E. Kennedy, U. Ozbek, and M. T. Dorak. What has GWAS done for HLA and disease associations? *International Journal of Immunogenetics*, 44(5):195–211, 2017. ISSN 1744313X. doi: 10.1111/iji.12332.

[63] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28:235–242, 2000. doi: 10.1093/nar/28.1.235.

[64] Pouya Faridi, Chen Li, Sri H. Ramarathinam, Patricia T. Illing, Nicole A. Mifsud, Rochelle Ayala, Jiangning Song, Linden J. Gearing, Nathan P. Croft, and Anthony W. Purcell. Response to Comment on "a subset of HLA-I peptides are not genomically templated: Evidence for cis- And trans-spliced peptide ligands". *Science Immunology*, 4(38), 2019. ISSN 24709468. doi: 10.1126/sciimmunol.aaw8457.

[65] Lu Deng, Ries J. Langley, Patrick H. Brown, Gang Xu, Leslie Teng, Qian Wang, Monica I. Gonzales, Glenda G. Callender, Michael I. Nishimura, Suzanne L. Topalian, and Roy A. Mariuzza. Structural basis for the recognition of mutant self by a tumor-specific, MHC class II-restricted T cell receptor. *Nature Immunology*, 8(4):398–408, 2007. ISSN 15292908. doi: 10.1038/ni1447.

[66] The PyMOL Molecular Graphics System, Version 2.4.2, 2021. URL `https://pymol.org/2/`.

[67] William Ferguson, Shira Dvora, Ronald W. Fikes, Anne C. Stone, and Stéphane Boissinot. Long-term balancing selection at the antiviral gene oas1 in central African chimpanzees. *Molecular Biology and Evolution*, 29(4):1093–1103, 2012. ISSN 07374038. doi: 10.1093/molbev/msr247.

[68] João C. Teixeira, Cesare De Filippo, Antje Weihmann, Juan R. Meneu, Fernando Racimo, Michael Dannemann, Birgit Nickel, Anne Fischer, Michel Halbwax, Claudine Andre, Rebeca Atencia, Matthias Meyer, Genís Parra, Svante Pääbo, and Aida M. Andrés. Long-term balancing selection in LAD1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. *Molecular Biology and Evolution*, 32(5):1186–1196, 2015. ISSN 15371719. doi: 10.1093/molbev/msv007.

[69] R. Cagliani, M. Fumagalli, M. Biasin, L. Piacentini, S. Riva, U. Pozzoli, M. C. Bonaglia, N. Bresolin, M. Clerici, and Manuela Sironi. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. *Human Genetics*, 128(6):577–588, 2010. ISSN 03406717. doi: 10.1007/s00439-010-0884-6.

[70] Ruchi M. Newman, Laura Hall, Michelle Connole, Guo Lin Chen, Shuji Sato, Eloisa Yuste, William Diehl, Eric Hunter, Amitinder Kaur, Gregory M. Miller, and Welkin E. Johnson. 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*, 103(50):19134–19139, 2006. ISSN 00278424. doi: 10.1073/pnas.0605838103.

[71] R. Cagliani, F. R. Guerini, M. Fumagalli, S. Riva, C. Agliardi, D. Galimberti, U. Pozzoli, A. Goris, B. Dubois, C. Fenoglio, D. Forni, S. Sanna, I. Zara, M. Pitzalis, M. Zoledziewska, F. Cucca, F. Marini, G. P. Comi, E. Scarpini, N. Bresolin, M. Clerici, and M. Sironi. A trans-specific polymorphism in ZC3HAV1 is maintained by long-standing balancing selection and may confer susceptibility

to multiple sclerosis. *Molecular Biology and Evolution*, 29(6):1599–1613, 2012. ISSN 07374038. doi: 10.1093/molbev/mss002.

[72] Shi Xia Xu, Wen Hua Ren, Shu Zhen Li, Fu Wen Wei, Kai Ya Zhou, and Guang Yang. Sequence polymorphism and evolution of three cetacean MHC genes. *Journal of Molecular Evolution*, 69(3): 260–275, 2009. ISSN 00222844. doi: 10.1007/s00239-009-9272-z.

[73] Li Li, Xiaopin Zhou, and Xiaolin Chen. Characterization and evolution of MHC class II B genes in ardeid birds. *Journal of Molecular Evolution*, 72(5):474–483, 2011. ISSN 00222844. doi: 10.1007/s00239-011-9446-3.

[74] Samit Kundu and Christopher G. Faulkes. A tangled history: Patterns of major histocompatibility complex evolution in the African mole-rats (Family: Bathyergidae). *Biological Journal of the Linnean Society*, 91(3):493–503, 2007. ISSN 00244066. doi: 10.1111/j.1095-8312.2007.00814.x.

[75] Mian Zhao, Yongzhen Wang, Hang Shen, Chenliang Li, Cheng Chen, Zhenhua Luo, and Hua Wu. Evolution by selection, recombination, and gene duplication in MHC class i genes of two Rhacophoridae species. *BMC Evolutionary Biology*, 13(1), 2013. ISSN 14712148. doi: 10.1186/1471-2148-13-113.

[76] Ikunari Kiryu, Johannes Martinus Dijkstra, Rafiqul Islam Sarder, Atsushi Fujiwara, Yasutoshi Yoshiura, and Mitsuru Ototake. New MHC class Ia domain lineages in rainbow trout (Oncorhynchus mykiss) which are shared with other fish species. *Fish and Shellfish Immunology*, 18(3):243–254, 2005. ISSN 10504648. doi: 10.1016/j.fsi.2004.07.007.

[77] Unni Grimholt, Kentaro Tsukamoto, Teruo Azuma, Jong Leong, Ben F. Koop, and Johannes M. Dijkstra. A comprehensive analysis of teleost MHC class I sequences. *BMC Evolutionary Biology*, 15 (1):1–17, 2015. ISSN 14712148. doi: 10.1186/s12862-015-0309-1.

[78] Emiko B. Sano, Christopher A. Wall, Patrick R. Hutchins, and Scott R. Miller. Ancient balancing selection on heterocyst function in a cosmopolitan cyanobacterium. *Nature Ecology and Evolution*, 2 (3):510–519, 2018. ISSN 2397334X. doi: 10.1038/s41559-017-0435-9.

[79] Thomas R. Ioerger, Andrew G. Clark, and Teh Hui Kao. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences of the United States of America*, 87(24):9732–9735, 1990. ISSN 00278424. doi: 10.1073/pnas.87.24.9732.

[80] Boris Igic, Lynn Bohs, and Joshua R. Kohn. Ancient polymorphism reveals unidirectional breeding system shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 103 (5):1359–1363, 2006. ISSN 00278424. doi: 10.1073/pnas.0506283103.

[81] Jessica Purcell, German Lagunas-Robles, Christian Rabeling, Marek L. Borowiec, and Alan Brelsford. The maintenance of polymorphism in an ancient social supergene. *Molecular Ecology*, 30(23): 6246–6258, 2021. ISSN 1365294X. doi: 10.1111/mec.16196.

[82] Débora Y.C. Brandt, Jônatas César, Jérôme Goudet, and Diogo Meyer. The effect of balancing selection on population differentiation: A study with HLA genes. *G3: Genes, Genomes, Genetics*, 8 (8):2805–2815, 2018. ISSN 21601836. doi: 10.1534/g3.118.200367.

[83] Kelly Nunes, Maria Helena Thomaz Maia, Eduardo José Melo dos Santos, Sidney Emanuel Batista dos Santos, João Farias Guerreiro, Maria Luiza Petzl-Erler, Gabriel Bedoya, Carla Gallo, Giovanni Poletti, Elena Llop, Luiza Tsuneto, Maria Cátira Bortolini, Francisco Rothhammer, Richard Single, Andrés Ruiz-Linares, Jorge Rocha, and Diogo Meyer. How natural selection shapes genetic differentiation in the MHC region: A case study with Native Americans. *Human Immunology*, 82(7):523–531, 2021. ISSN 18791166. doi: 10.1016/j.humimm.2021.03.005.

[84] Gaurav Bhatia, Nick Patterson, Bogdan Pasaniuc, Noah Zaitlen, Giulio Genovese, Samuela Pollack, Swapan Mallick, Simon Myers, Arti Tandon, Chris Spencer, Cameron D. Palmer, Adebowale A. Adeyemo, Ermeg L. Akylbekova, L. Adrienne Cupples, Jasmin Divers, Myriam Fornage, W. H.Linda

Kao, Leslie Lange, Mingyao Li, Solomon Musani, Josyf C. Mychaleckyj, Adesola Ogunniyi, George Papanicolaou, Charles N. Rotimi, Jerome I. Rotter, Ingo Ruczinski, Babatunde Salako, David S. Siscovick, Bamidele O. Tayo, Qiong Yang, Steve McCarroll, Pardis Sabeti, Guillaume Lettre, Phil De Jager, Joel Hirschhorn, Xiaofeng Zhu, Richard Cooper, David Reich, James G. Wilson, and Alkes L. Price. Genome-wide comparison of African-ancestry populations from CARe and other cohorts reveals signals of natural selection. *American Journal of Human Genetics*, 89(3):368–381, 2011. ISSN 00029297. doi: 10.1016/j.ajhg.2011.07.025.

[85] Juan Antonio Vizcaíno, Peter Kubiniok, Kevin A. Kovalchik, Qing Ma, Jérôme D. Duquette, Ian Mongrain, Eric W. Deutsch, Bjoern Peters, Alessandro Sette, Isabelle Sirois, and Etienne Caron. The human immunopeptidome project: A roadmap to predict and treat immune diseases. *Molecular and Cellular Proteomics*, 19(1):31–49, 2020. ISSN 15359484. doi: 10.1074/mcp.R119.001743.

[86] Robert C. Edgar. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004. ISSN 03051048. doi: 10.1093/nar/gkh340.

[87] Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35 (6):1547–1549, 2018. ISSN 15371719. doi: 10.1093/molbev/msy096.

[88] The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526 (7571):68–74, 2015. ISSN 14764687. doi: 10.1038/nature15393.

[89] 3,202 samples at high-coverage from NYGC, 2020. URL `https://www.internationalgenome.org/announcements/3202-samples-at-high-coverage-from-NYGC/`.

[90] Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, and Richard Durbin. The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158, 2011. ISSN 13674803. doi: 10.1093/bioinformatics/btr330.

[91] Andrew Rambaut, Alexei J. Drummond, Dong Xie, Guy Baele, and Marc A. Suchard. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5):901–904, 2018. ISSN 1076836X. doi: 10.1093/sysbio/syy032.

[92] Alexei J. Drummond and Andrew Rambaut. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, 7(1):1–8, 2007. ISSN 14712148. doi: 10.1186/1471-2148-7-214.

[93] Nicola F. Müller and Remco R. Bouckaert. Adaptive metropolis-coupled MCMC for BEAST 2. *PeerJ*, 8:1–16, 2020. ISSN 21678359. doi: 10.7717/peerj.9473.

[94] Jun Liu, Shihong Zhang, Shuguang Tan, Yong Yi, Bin Wu, Bin Cao, Fengcai Zhu, Chen Wang, Hua Wang, Jianxun Qi, and George F. Gao. Cross-Allele Cytotoxic T Lymphocyte Responses against 2009 Pandemic H1N1 Influenza A Virus among HLA-A24 and HLA-A3 Supertype-Positive Individuals. *Journal of Virology*, 86(24):13281–13294, 2012. ISSN 0022-538X. doi: 10.1128/jvi.01841-12.

[95] Joanna A. L. Choo, Jingxian Liu, Xinyu Toh, Gijsbert M. Grotenbreg, and Ee Chee Ren. 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*, 88(18): 10613–10623, 2014. ISSN 0022-538X. doi: 10.1128/jvi.00855-14.

[96] Gina M. Clayton, Yang Wang, Frances Crawford, Andrey Novikov, Brian T. Wimberly, Jeffrey S. Kieft, Michael T. Falta, Natalie A. Bowerman, Philippa Marrack, Andrew P. Fontenot, Shaodong Dai, and John W. Kappler. Structural basis of chronic beryllium disease: Linking allergic hypersensitivity and autoimmunity. *Cell*, 158(1):132–142, 2014. ISSN 10974172. doi: 10.1016/j.cell.2014.04.048.

[97] Christian Siebold, Bjarke E. Hansen, Jessica R. Wyer, Karl Harlos, Robert E. Esnouf, Arne Svejgaard, John I. Bell, Jack L. Strominger, E. Yvonne Jones, and Lars Fugger. Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy.

Proceedings of the National Academy of Sciences of the United States of America, 101(7):1999–2004, 2004. ISSN 00278424. doi: 10.1073/pnas.0308458100.

[98] Shiyan Zhu, Kefang Liu, Yan Chai, Yanan Wu, Dan Lu, Wenling Xiao, Hao Cheng, Yingze Zhao, Chunming Ding, Jianxin Lyu, Yongliang Lou, George F. Gao, and William J. Liu. Divergent Peptide Presentations of HLA-A*30 Alleles Revealed by Structures With Pathogen Peptides. *Frontiers in immunology*, 10:1709, 2019. ISSN 16643224. doi: 10.3389/fimmu.2019.01709.

[99] Yeping Sun, Jun Liu, Meng Yang, Feng Gao, Jianfang Zhou, Yoshihiro Kitamura, Bin Gao, Po Tien, Yuelong Shu, Aikichi Iwamoto, Zhu Chen, and George F. Gao. Identification and structural definition of H5-specific CTL epitopes restricted by HLA-A*0201 derived from the H5N1 subtype of influenza A viruses. *Journal of General Virology*, 91(4):919–930, 2010. ISSN 00221317. doi: 10.1099/vir.0.016766-0.

[100] Pravin Kumar, Ardeschir Vahedi-Faridi, Wolfram Saenger, Elena Merino, José A. López de Castro, Barbara Uchanska-Ziegler, and Andreas Ziegler. Structural basis for T cell alloreactivity among three HLA-B14 and HLA-B27 antigens. *Journal of Biological Chemistry*, 284(43):29784–29797, 2009. ISSN 00219258. doi: 10.1074/jbc.M109.038497.

[101] Terry C.C. Lim Kam Sian, Saranjah Indumathy, Hanim Halim, Anja Greule, Max J. Cryle, Paul Bowness, Jamie Rossjohn, Stephanie Gras, Anthony W. Purcell, and Ralf B. Schittenhelm. Allelic association with ankylosing spondylitis fails to correlate with human leukocyte antigen b27 homodimer formation. *Journal of Biological Chemistry*, 294(52):20185–20195, 2019. ISSN 1083351X. doi: 10.1074/jbc.RA119.010257.

[102] Jesse I. Mobbs, Patricia T. Illing, Nadine L. Dudek, Andrew G. Brooks, Daniel G. Baker, Anthony W. Purcell, Jamie Rossjohn, and Julian P. Vivian. The molecular basis for peptide repertoire selection in the human leucocyte antigen (HLA) C06:02 molecule. *Journal of Biological Chemistry*, 292(42):17203–17215, 2017. ISSN 1083351X. doi: 10.1074/jbc.M117.806976.

[103] Gurman Kaur, Stephanie Gras, Jesse I. Mobbs, Julian P. Vivian, Adrian Cortes, Thomas Barber, Subita Balaram Kuttikkatte, Lise Torp Jensen, Kathrine E. Attfield, Calliope A. Dendrou, Mary Carrington, Gil Mcvean, Anthony W. Purcell, Jamie Rossjohn, and Lars Fugger. Structural and regulatory diversity shape HLA-C protein expression levels. *Nature Communications*, 8(1):1–12, 2017. ISSN 20411723. doi: 10.1038/ncomms15924.

[104] Shaodong Dai, Guinevere A. Murphy, Frances Crawford, Douglas G. Mack, Michael T. Falta, Philippa Marrack, John W. Kappler, and Andrew P. Fontenot. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7425–7430, 2010. ISSN 00278424. doi: 10.1073/pnas.1001772107.

[105] Seisuke Kusano, Mutsuko Kukimoto-Niino, Yoko Satta, Noboru Ohsawa, Tomomi Uchikubo-Kamo, Motoaki Wakiyama, Mariko Ikeda, Takaho Terada, Ken Yamamoto, Yasuharu Nishimura, Mikako Shirouzu, Takehiko Sasazuki, and Shigeyuki Yokoyama. 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*, 426(17):3016–3027, 2014. ISSN 10898638. doi: 10.1016/j.jmb.2014.06.020.

[106] Kate N. Henderson, Jason A. Tye-Din, Hugh H. Reid, Zhenjun Chen, Natalie A. Borg, Tim Beissbarth, Arthur Tatham, Stuart I. Mannering, Anthony W. Purcell, Nadine L. Dudek, David A. van Heel, James McCluskey, Jamie Rossjohn, and Robert P. Anderson. A Structural and Immunological Basis for the Role of Human Leukocyte Antigen DQ8 in Celiac Disease. *Immunity*, 27(1):23–34, 2007. ISSN 10747613. doi: 10.1016/j.immuni.2007.05.015.

[107] Stig Tollefsen, Kinya Hotta, Xi Chen, Bjørg Simonsen, Kunchithapadam Swaminathan, Irimpan I. Mathews, Ludvig M. Sollid, and Chu Young Kim. Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. *Journal of Biological Chemistry*, 287(17): 13611–13619, 2012. ISSN 00219258. doi: 10.1074/jbc.M111.320374.

[108] Stephen W. Scally, Soi Cheng Law, Yi Tian Ting, Jurgen Van Heemst, Jeremy Sokolove, Aaron J. Deutsch, E. Bridie Clemens, Antonis K. Moustakas, George K. Papadopoulos, Diane Van Der Woude, Irene Smolik, Carol A. Hitchon, David B. Robinson, Elizabeth D. Ferucci, Charles N. Bernstein, Xiaobo Meng, Vidyanand Anaparti, Tom Huizinga, Katherine Kedzierska, Hugh H. Reid, Soumya Raychaudhuri, René E. Toes, Jamie Rossjohn, Hani El-Gabalawy, and Ranjeny Thomas. Molecular basis for increased susceptibility of Indigenous North Americans to seropositive rheumatoid arthritis. *Annals of the Rheumatic Diseases*, 76(11):1915–1923, 2017. ISSN 14682060. doi: 10.1136/annrheumdis-2017-211300.

[109] Christina Gerstner, Anatoly Dubnovitsky, Charlotta Sandin, Genadiy Kozhukh, Hannes Uchtenhagen, Eddie A. James, Johan Rönnelid, Anders Jimmy Ytterberg, Jennifer Pieper, Evan Reed, Carolina Tandre, Mary Rieck, Roman A. Zubarev, Lars Rönnblom, Tatyana Sandalova, Jane H. Buckner, Adnane Achour, and Vivianne Malmström. Functional and structural characterization of a novel HLA-DRB1*04: 01-restricted α-enolase T cell epitope in rheumatoid arthritis. *Frontiers in Immunology*, 7:1–12, 2016. ISSN 16643224. doi: 10.3389/fimmu.2016.00494.

[110] Yang Luo, Masahiro Kanai, Wanson Choi, Xinyi Li, Saori Sakaue, Kenichi Yamamoto, Kotaro Ogawa, Maria Gutierrez-Arcelus, Peter K. Gregersen, Philip E. Stuart, James T. Elder, Lukas Forer, Sebastian Schönherr, Christian Fuchsberger, Albert V. Smith, Jacques Fellay, Mary Carrington, David W. Haas, Xiuqing Guo, Nicholette D. Palmer, Yii Der Ida Chen, Jerome I. Rotter, Kent D. Taylor, Stephen S. Rich, Adolfo Correa, James G. Wilson, Sekar Kathiresan, Michael H. Cho, Andres Metspalu, Tonu Esko, Yukinori Okada, Buhm Han, Namiko Abe, Gonçalo Abecasis, Francois Aguet, Christine Albert, Laura Almasy, Alvaro Alonso, Seth Ament, Peter Anderson, Pramod Anugu, Deborah Applebaum-Bowden, Kristin Ardlie, Dan Arking, Donna K. Arnett, Allison Ashley-Koch, Stella Aslibekyan, Tim Assimes, Paul Auer, Dimitrios Avramopoulos, Najib Ayas, Adithya Balasubramanian, John Barnard, Kathleen Barnes, R. Graham Barr, Emily Barron-Casella, Lucas Barwick, Terri Beaty, Gerald Beck, Diane Becker, Lewis Becker, Rebecca Beer, Amber Beitelshees, Emelia Benjamin, Takis Benos, Marcos Bezerra, Larry Bielak, Joshua Bis, Thomas Blackwell, John Blangero, Eric Boerwinkle, Donald W. Bowden, Russell Bowler, Jennifer Brody, Ulrich Broeckel, Jai Broome, Deborah Brown, Karen Bunting, Esteban Burchard, Carlos Bustamante, Erin Buth, Brian Cade, Jonathan Cardwell, Vincent Carey, Julie Carrier, Cara Carty, Richard Casaburi, Juan P.Casas Romero, James Casella, Peter Castaldi, Mark Chaffin, Christy Chang, Yi Cheng Chang, Daniel Chasman, Sameer Chavan, Bo Juen Chen, Wei Min Chen, Seung Hoan Choi, Lee Ming Chuang, Mina Chung, Ren Hua Chung, Clary Clish, Suzy Comhair, Matthew Conomos, Elaine Cornell, Carolyn Crandall, James Crapo, L. Adrienne Cupples, Joanne Curran, Jeffrey Curtis, Brian Custer, Coleen Damcott, Dawood Darbar, Sean David, Colleen Davis, Michelle Daya, Mariza de Andrade, Lisa de las Fuentes, Paul de Vries, Michael DeBaun, Ranjan Deka, Dawn DeMeo, Scott Devine, Huyen Dinh, Harsha Doddapaneni, Qing Duan, Shannon Dugan-Perez, Ravi Duggirala, Jon Peter Durda, Susan K. Dutcher, Charles Eaton, Lynette Ekunwe, Adel El Boueiz, Patrick Ellinor, Leslie Emery, Serpil Erzurum, Charles Farber, Jesse Farek, Tasha Fingerlin, Matthew Flickinger, Myriam Fornage, Nora Franceschini, Chris Frazar, Mao Fu, Stephanie M. Fullerton, Lucinda Fulton, Stacey Gabriel, Weiniu Gan, Shanshan Gao, Yan Gao, Margery Gass, Heather Geiger, Bruce Gelb, Mark Geraci, Soren Germer, Robert Gerszten, Auyon Ghosh, Richard Gibbs, Chris Gignoux, Mark Gladwin, David Glahn, Stephanie Gogarten, Da Wei Gong, Harald Goring, Sharon Graw, Kathryn J. Gray, Daniel Grine, Colin Gross, C. Charles Gu, Yue Guan, Namrata Gupta, David M. Haas, Jeff Haessler, Michael Hall, Yi Han, Patrick Hanly, Daniel Harris, Nicola L. Hawley, Jiang He, Ben Heavner, Susan Heckbert, Ryan Hernandez, David Herrington, Craig Hersh, Bertha Hidalgo, James Hixson, Brian Hobbs, John Hokanson, Elliott Hong, Karin Hoth, Chao (Agnes) Hsiung, Jianhong Hu, Yi Jen Hung, Haley Huston, Chii Min Hwu, Marguerite Ryan Irvin, Rebecca Jackson, Deepti Jain, Cashell Jaquish, Jill Johnsen, Andrew Johnson, Craig Johnson, Rich Johnston, Kimberly Jones, Hyun Min Kang, Robert Kaplan, Sharon Kardia, Shannon Kelly, Eimear Kenny, Michael Kessler, Alyna Khan, Ziad Khan, Wonji Kim, John Kimoff, Greg Kinney, Barbara Konkle, Charles Kooperberg, Holly Kramer, Christoph Lange, Ethan Lange, Leslie Lange, Cathy Laurie, Cecelia Laurie, Meryl LeBoff, Jiwon Lee, Sandra Lee, Wen Jane Lee, Jonathon LeFaive, David Levine, Dan Levy, Joshua Lewis, Xiaohui Li, Yun Li, Henry Lin, Honghuang Lin, Xihong Lin, Simin Liu,

Yongmei Liu, Yu Liu, Ruth J.F. Loos, Steven Lubitz, Kathryn Lunetta, James Luo, Ulysses Magalang, Michael Mahaney, Barry Make, Ani Manichaikul, Alisa Manning, Jo Ann Manson, Lisa Martin, Melissa Marton, Susan Mathai, Rasika Mathias, Susanne May, Patrick McArdle, Merry Lynn McDonald, Sean McFarland, Stephen McGarvey, Daniel McGoldrick, Caitlin McHugh, Becky McNeil, Hao Mei, James Meigs, Vipin Menon, Luisa Mestroni, Ginger Metcalf, Deborah A. Meyers, Emmanuel Mignot, Julie Mikulla, Nancy Min, Mollie Minear, Ryan L. Minster, Braxton D. Mitchell, Matt Moll, Zeineen Momin, May E. Montasser, Courtney Montgomery, Donna Muzny, Josyf C. Mychaleckyj, Girish Nadkarni, Rakhi Naik, Take Naseri, Pradeep Natarajan, Sergei Nekhai, Sarah C. Nelson, Bonnie Neltner, Caitlin Nessner, Deborah Nickerson, Osuji Nkechinyere, Kari North, Jeff O'Connell, Tim O'Connor, Heather Ochs-Balcom, Geoffrey Okwuonu, Allan Pack, David T. Paik, James Pankow, George Papanicolaou, Cora Parker, Gina Peloso, Juan Manuel Peralta, Marco Perez, James Perry, Ulrike Peters, Patricia Peyser, Lawrence S. Phillips, Jacob Pleiness, Toni Pollin, Wendy Post, Julia Powers Becker, Meher Preethi Boorgula, Michael Preuss, Bruce Psaty, Pankaj Qasba, Dandi Qiao, Zhaohui Qin, Nicholas Rafaels, Laura Raffield, Mahitha Rajendran, Vasan S. Ramachandran, D. C. Rao, Laura Rasmussen-Torvik, Aakrosh Ratan, Susan Redline, Robert Reed, Catherine Reeves, Elizabeth Regan, Alex Reiner, Muagututi'a Sefuiva Reupena, Ken Rice, Rebecca Robillard, Nicolas Robine, Dan Roden, Carolina Roselli, Ingo Ruczinski, Alexi Runnels, Pamela Russell, Sarah Ruuska, Kathleen Ryan, Ester Cerdeira Sabino, Danish Saleheen, Shabnam Salimi, Sejal Salvi, Steven Salzberg, Kevin Sandow, Vijay G. Sankaran, Jireh Santibanez, Karen Schwander, David Schwartz, Frank Sciurba, Christine Seidman, Jonathan Seidman, Frédéric Sériès, Vivien Sheehan, Stephanie L. Sherman, Amol Shetty, Aniket Shetty, Wayne Hui Heng Sheu, M. Benjamin Shoemaker, Brian Silver, Edwin Silverman, Robert Skomro, Jennifer Smith, Josh Smith, Nicholas Smith, Tanja Smith, Sylvia Smoller, Beverly Snively, Michael Snyder, Tamar Sofer, Nona Sotoodehnia, Adrienne M. Stilp, Garrett Storm, Elizabeth Streeten, Jessica Lasky Su, Yun Ju Sung, Jody Sylvia, Adam Szpiro, Daniel Taliun, Hua Tang, Margaret Taub, Matthew Taylor, Simeon Taylor, Marilyn Telen, Timothy A. Thornton, Machiko Threlkeld, Lesley Tinker, David Tirschwell, Sarah Tishkoff, Hemant Tiwari, Catherine Tong, Russell Tracy, Michael Tsai, Dhananjay Vaidya, David Van Den Berg, Peter VandeHaar, Scott Vrieze, Tarik Walker, Robert Wallace, Avram Walts, Fei Fei Wang, Heming Wang, Jiongming Wang, Karol Watson, Jennifer Watt, Daniel E. Weeks, Joshua Weinstock, Bruce Weir, Scott T. Weiss, Lu Chen Weng, Jennifer Wessel, Cristen Willer, Kayleen Williams, L. Keoki Williams, Carla Wilson, James Wilson, Lara Winterkorn, Quenna Wong, Joseph Wu, Huichun Xu, Lisa Yanek, Ivana Yang, Ketian Yu, Seyedeh Maryam Zekavat, Yingze Zhang, Snow Xueyan Zhao, Wei Zhao, Xiaofeng Zhu, Michael Zody, Sebastian Zoellner, Paul J. McLaren, and Soumya Raychaudhuri. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nature Genetics*, 53(10):1504–1516, 2021. ISSN 15461718. doi: 10.1038/s41588-021-00935-7.

[111] Mao Fang Luo and Hui Juan Pan. MHC II DRB variation and trans-species polymorphism in the golden snub-nosed monkey (Rhinopithecus roxellana). *Chinese Science Bulletin*, 58(18):2119–2127, 2013. ISSN 10016538. doi: 10.1007/s11434-013-5713-6.

[112] Angela Román, Isabel Cervera, Jacqueline Head, Miriam Rodríguez, and Jorge Martínez-Laso. Generation of HLA-B*1516/B*1567/B*1595 and B*1517 alleles (B15 specific group) by transpecies evolution. *Human Immunology*, 68(12):1001–1008, 2007. ISSN 01988859. doi: 10.1016/j.humimm.2007.10.010.

[113] Y. Cao, A. Li, L. Li, X. Yan, Y. Fa, L. Zeng, J. Fan, B. Liu, and Z. Sun. Identification of 32 major histocompatibility complex class I alleles in African green monkeys. *Tissue Antigens*, 84(3):304–307, 2014. ISSN 13990039. doi: 10.1111/tan.12389.

[114] L. K. Gaur, G. T. Nepom, K. E. Snyder, J. Anderson, M. Pandarpurkar, W. Yadock, and E. R. Heise. MHC-DRB allelic sequences incorporate distinct intragenic trans-specific segments. *Tissue Antigens*, 49(4):342–355, 1997. ISSN 00012815. doi: 10.1111/j.1399-0039.1997.tb02762.x.

[115] Pei Zhang, Kang Huang, Bingyi Zhang, Derek W. Dunn, Dan Chen, Fan Li, Xiaoguang Qi, Songtao Guo, and Baoguo Li. High polymorphism in MHC-DRB genes in golden snub-nosed monkeys re-

veals balancing selection in small, isolated populations. *BMC Evolutionary Biology*, 18(1):1–13, 2018. ISSN 14712148. doi: 10.1186/s12862-018-1148-7.

[116] K. Sano, T. Shiina, S. Kohara, K. Yanagiya, K. Hosomichi, S. Shimizu, T. Anzai, A. Watanabe, K. Ogasawara, R. Torii, J. K. Kulski, and H. Inoko. Novel cynomolgus macaque MHC-DPB1 polymorphisms in three South-East Asian populations. *Tissue Antigens*, 67(4):297–306, 2006. ISSN 00012815. doi: 10.1111/j.1399-0039.2006.00577.x.

[117] Y. F. Yao, J. J. Zhao, Q. X. Dai, J. Y. Li, L. Zhou, Y. T. Wang, Q. Y. Ni, M. W. Zhang, and H. L. Xu. Identification and characterization of the major histocompatibility complex class II DQB (MhcMath-DQB1) alleles in Tibetan macaques (Macaca thibetana). *Tissue Antigens*, 82(2):113–121, 2013. ISSN 00012815. doi: 10.1111/tan.12145.

[118] Yong Fang Yao, Qiu Xia Dai, Jing Li, Qing Yong Ni, Ming Wang Zhang, and Huai Liang Xu. Genetic diversity and differentiation of the rhesus macaque (Macaca mulatta) population in western Sichuan, China, based on the second exon of the major histocompatibility complex class II DQB (MhcMamu-DQB1) alleles. *BMC Evolutionary Biology*, 14(1):1–13, 2014. ISSN 14712148. doi: 10.1186/1471-2148-14-130.

[119] Yoshiki Yasukochi and Yoko Satta. A human-specific allelic group of the MHC DRB1 gene in primates. *Journal of Physiological Anthropology*, 33(1):1–9, 2014. ISSN 18806805. doi: 10.1186/1880-6805-33-14.

[120] Bastiaan L. Slierendregt, Nel Otting, Nicole Van Besouw, Margreet Jonker, and Ronald E. Bontrop. Expansion and contraction of rhesus macaque DRB regions by duplication and deletion. *Journal of Immunology*, 152(5):2298–2307, 1994. ISSN 0022-1767.

[121] C. F. Suárez, P. P. Cárdenas, E. J. Llanos-Ballestas, P. Martínez, M. Obregón, M. E. Patarroyo, and M. A. Patarroyo. α1 and α2 domains of Aotus MHC class I and catarrhini MHC class Ia share similar characteristics. *Tissue Antigens*, 61(5):362–373, 2003. ISSN 00012815. doi: 10.1034/j.1399-0039.2003.00045.x.

[122] Janet C. Buckner, Katharine M. Jack, Amanda D. Melin, Valérie A.M. Schoof, Gustavo A. Gutiérrez-Espeleta, Marcela G.M. Lima, and Jessica W. Lynch. Major histocompatibility complex class II DR and DQ evolution and variation in wild capuchin monkey species (Cebinae). *PLoS ONE*, 16(8):e0254604, 2021. ISSN 19326203. doi: 10.1371/journal.pone.0254604.

[123] John P. Huelsenbeck and Fredrik Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8):754–755, 2001. ISSN 13674803. doi: 10.1093/bioinformatics/17.8.754.

[124] Xiaoyue Song, Pei Zhang, Kang Huang, Dan Chen, Songtao Guo, Xiaoguang Qi, Gang He, Ruliang Pan, and Baoguo Li. The influence of positive selection and trans-species evolution on DPB diversity in the golden snub-nosed monkeys (Rhinopithecus roxellana). *Primates*, 57(4):489–499, 2016. ISSN 00328332. doi: 10.1007/s10329-016-0544-0.

[125] Dagan A. Loisel, Matthew V. Rockman, Gregory A. Wray, Jeanne Altmann, and Susan C. Alberts. 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*, 103(44):16331–16336, 2006. ISSN 00278424. doi: 10.1073/pnas.0607662103.

[126] Noah D. Simons, Geeta N. Eick, Maria J. Ruiz-Lopez, Patrick A. Omeja, Colin A. Chapman, Tony L. Goldberg, Nelson Ting, and Kirstin N. Sterner. Cis-regulatory evolution in a wild primate: Infection-associated genetic variation drives differential expression of MHC-DQA1 in vitro. *Molecular Ecology*, 26(17):4523–4535, 2017. ISSN 1365294X. doi: 10.1111/mec.14221.

[127] María José Villanueva-Noriega, Charles Scott Baker, and Luis Medrano-González. Evolution of the MHC-DQB exon 2 in marine and terrestrial mammals. *Immunogenetics*, 65(1):47–61, 2013. ISSN 00937711. doi: 10.1007/s00251-012-0647-8.

[128] Nick Patterson, Daniel J. Richter, Sante Gnerre, Eric S. Lander, and David Reich. Genetic evidence for complex speciation of humans and chimpanzees. *Nature*, 441(7097):1103–1108, 2006. ISSN 14764687. doi: 10.1038/nature04789.

[129] Aylwyn Scally, Julien Y. Dutheil, Ladeana W. Hillier, Gregory E. Jordan, Ian Goodhead, Javier Herrero, Asger Hobolth, Tuuli Lappalainen, Thomas Mailund, Tomas Marques-Bonet, Shane McCarthy, Stephen H. Montgomery, Petra C. Schwalie, Y. Amy Tang, Michelle C. Ward, Yali Xue, Bryndis Yngvadottir, Can Alkan, Lars N. Andersen, Qasim Ayub, Edward V. Ball, Kathryn Beal, Brenda J. Bradley, Yuan Chen, Chris M. Clee, Stephen Fitzgerald, Tina A. Graves, Yong Gu, Paul Heath, Andreas Heger, Emre Karakoc, Anja Kolb-Kokocinski, Gavin K. Laird, Gerton Lunter, Stephen Meader, Matthew Mort, James C. Mullikin, Kasper Munch, Timothy D. O'Connor, Andrew D. Phillips, Javier Prado-Martinez, Anthony S. Rogers, Saba Sajjadian, Dominic Schmidt, Katy Shaw, Jared T. Simpson, Peter D. Stenson, Daniel J. Turner, Linda Vigilant, Albert J. Vilella, Weldon Whitener, Baoli Zhu, David N. Cooper, Pieter De Jong, Emmanouil T. Dermitzakis, Evan E. Eichler, Paul Flicek, Nick Goldman, Nicholas I. Mundy, Zemin Ning, Duncan T. Odom, Chris P. Ponting, Michael A. Quail, Oliver A. Ryder, Stephen M. Searle, Wesley C. Warren, Richard K. Wilson, Mikkel H. Schierup, Jane Rogers, Chris Tyler-Smith, and Richard Durbin. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 483(7388):169–175, 2012. ISSN 00280836. doi: 10.1038/nature10842.

[130] Knut Finstermeier, Dietmar Zinner, Markus Brameier, Matthias Meyer, Eva Kreuz, Michael Hofreiter, and Christian Roos. A Mitogenomic Phylogeny of Living Primates. *PLoS ONE*, 8(7):1–10, 2013. ISSN 19326203. doi: 10.1371/journal.pone.0069504.

[131] James E. Tarver, Mario Dos Reis, Siavash Mirarab, Raymond J. Moran, Sean Parker, Joseph E. O'Reilly, Benjamin L. King, Mary J. O'Connell, Robert J. Asher, Tandy Warnow, Kevin J. Peterson, Philip C.J. Donoghue, and Davide Pisani. The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome Biology and Evolution*, 8(2):330–344, 2016. ISSN 17596653. doi: 10.1093/gbe/evv261.

[132] Michael J. Benton, Philip C.J. Donoghue, Robert J. Asher, Matt Friedman, Thomas J. Near, and Jakob Vinther. Constraints on the timescale of animal evolutionary history. *Palaeontologia Electronica*, 18(1):1–107, 2015. ISSN 10948074. doi: 10.26879/424.

[133] Christian Roos and Dietmar Zinner. Primate Phylogeny. *The International Encyclopedia of Primatology*, pages 1–5, 2017. doi: 10.1002/9781119179313.wbprim0394.

[134] Søren Besenbacher, Christina Hvilsom, Tomas Marques-Bonet, Thomas Mailund, and Mikkel Heide Schierup. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nature Ecology and Evolution*, 3(2):286–292, 2019. ISSN 2397334X. doi: 10.1038/s41559-018-0778-x.

[135] Dan Vanderpool, Bui Quang Minh, Robert Lanfear, Daniel Hughes, Shwetha Murali, R. Alan Harris, Muthuswamy Raveendran, Donna M. Muzny, Mark S. Hibbins, Robert J. Williamson, Richard A. Gibbs, Kim C. Worley, Jeffrey Rogers, and Matthew W. Hahn. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *PLoS Biology*, 18(12):1–27, 2020. ISSN 15457885. doi: 10.1371/journal.pbio.3000954.

[136] Sergio Almécija, Ashley S. Hammond, Nathan E. Thompson, Kelsey D. Pugh, Salvador Moyà-Solà, and David M. Alba. Fossil apes and human evolution. *Science*, 372(6542), 2021. ISSN 10959203. doi: 10.1126/science.abb4363.

[137] Mario Dos Reis, Gregg F. Gunnell, Jose Barba-Montoya, Alex Wilkins, Ziheng Yang, and Anne D. Yoder. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: Primates as a test case. *Systematic Biology*, 67(4):594–615, 2018. ISSN 1076836X. doi: 10.1093/sysbio/syy001.

[138] Polina Perelman, Warren E. Johnson, Christian Roos, Hector N. Seuánez, Julie E. Horvath, Miguel A.M. Moreira, Bailey Kessing, Joan Pontius, Melody Roelke, Yves Rumpler, Maria Paula C. Schneider, Artur Silva, Stephen J. O'Brien, and Jill Pecon-Slattery. A molecular phylogeny of living primates. *PLoS Genetics*, 7(3):1–17, 2011. ISSN 15537404. doi: 10.1371/journal.pgen.1001342.

[139] Iyad S. Zalmout, William J. Sanders, Laura M. MacLatchy, Gregg F. Gunnell, Yahya A. Al-Mufarreh, Mohammad A. Ali, Abdul Azziz H. Nasser, Abdu M. Al-Masari, Salih A. Al-Sobhi, Ayman O. Nadhra, Adel H. Matari, Jeffrey A. Wilson, and Philip D. Gingerich. New oligocene primate from Saudi Arabia and the divergence of apes and old world monkeys. *Nature*, 466(7304):360–364, 2010. ISSN 00280836. doi: 10.1038/nature09094.

[140] Nancy J. Stevens, Erik R. Seiffert, Patrick M. O'Connor, Eric M. Roberts, Mark D. Schmitz, Cornelia Krause, Eric Gorscak, Sifa Ngasala, Tobin L. Hieronymus, and Joseph Temu. Palaeontological evidence for an Oligocene divergence between Old World monkeys and apes. *Nature*, 497(7451):611–614, 2013. ISSN 00280836. doi: 10.1038/nature12161.

[141] Nathan S. Upham, Jacob A. Esselstyn, and Walter Jetz. Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, 17(12):1–44, 2019. ISSN 15457885. doi: 10.1371/journal.pbio.3000494.

[142] Juan C. Opazo, Derek E. Wildman, Tom Prychitko, Robert M. Johnson, and Morris Goodman. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). *Molecular Phylogenetics and Evolution*, 40(1):274–280, 2006. ISSN 10557903. doi: 10.1016/j.ympev.2005.11.015.

[143] Richard F. Kay. New World monkey origins. *Science*, 347(6226):1068–1069, 2015. ISSN 10959203. doi: 10.1126/science.aaa9217.

[144] Maureen a O Leary, Jonathan I Bloch, John J Flynn, Timothy J Gaudin, Andres Giallombardo, Norberto P Giannini, Suzann L Goldberg, Brian P Kraatz, Zhe-xi Luo, Jin Meng, Xijun Ni, Michael J Novacek, Fernando a Perini, Zachary S Randall, Guillermo W Rougier, Eric J Sargis, Mary T Silcox, Nancy B Simmons, Michelle Spaulding, Paúl M Velazco, Marcelo Weksler, John R Wible, and Andrea L Cirranello. The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science*, 339(6120):662–667, 2013. doi: 10.1126/science.1229237.

[145] Chao Yuan, Hao Chen, and Daisuke Kihara. Effective inter-residue contact definitions for accurate protein fold recognition. *BMC Bioinformatics*, 13(1), 2012. ISSN 14712105. doi: 10.1186/1471-2105-13-292.

[146] Chao Tian, Bethann S. Hromatka, Amy K. Kiefer, Nicholas Eriksson, Suzanne M. Noble, Joyce Y. Tung, and David A. Hinds. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nature Communications*, 8(1), 2017. ISSN 20411723. doi: 10.1038/s41467-017-00257-5.

[147] Saori Sakaue, Masahiro Kanai, Yosuke Tanigawa, Juha Karjalainen, Mitja Kurki, Seizo Koshiba, Akira Narita, Takahiro Konuma, Kenichi Yamamoto, Masato Akiyama, Kazuyoshi Ishigaki, Akari Suzuki, Ken Suzuki, Wataru Obara, Ken Yamaji, Kazuhisa Takahashi, Satoshi Asai, Yasuo Takahashi, Takao Suzuki, Nobuaki Shinozaki, Hiroki Yamaguchi, Shiro Minami, Shigeo Murayama, Kozo Yoshimori, Satoshi Nagayama, Daisuke Obata, Masahiko Higashiyama, Akihide Masumoto, Yukihiro Koretsune, Kaoru Ito, Chikashi Terao, Toshimasa Yamauchi, Issei Komuro, Takashi Kadowaki, Gen Tamiya, Masayuki Yamamoto, Yusuke Nakamura, Michiaki Kubo, Yoshinori Murakami, Kazuhiko Yamamoto, Yoichiro Kamatani, Aarno Palotie, Manuel A. Rivas, Mark J. Daly, Koichi Matsuda, and Yukinori Okada. A cross-population atlas of genetic associations for 220 human phenotypes. *Nature Genetics*, 53(10):1415–1424, 2021. ISSN 15461718. doi: 10.1038/s41588-021-00931-x.

[148] Sarah J. Dunstan, Nguyen Thi Hue, Buhm Han, Zheng Li, Trinh Thi Bich Tram, Kar Seng Sim, Christopher M. Parry, Nguyen Tran Chinh, Ha Vinh, Nguyen Phu Huong Lan, Nga Tran Vu Thieu, Phat Voong Vinh, Samir Koirala, Sabina Dongol, Amit Arjyal, Abhilasha Karkey, Olita Shilpakar,

Christiane Dolecek, Jia Nee Foo, Le Thi Phuong, Mai Ngoc Lanh, Tan Do, Tin Aung, Do Nu Hon, Yik Ying Teo, Martin L. Hibberd, Katherine L. Anders, Yukinori Okada, Soumya Raychaudhuri, Cameron P. Simmons, Stephen Baker, Paul I.W. De Bakker, Buddha Basnyat, Tran Tinh Hien, Jeremy J. Farrar, and Chiea Chuen Khor. Variation at HLA-DRB1 is associated with resistance to enteric fever. *Nature Genetics*, 46(12):1333–1336, 2014. ISSN 15461718. doi: 10.1038/ng.3143.

[149] Zhenzhen Wang, Yonghu Sun, Xi'An Fu, Gongqi Yu, Chuan Wang, Fangfang Bao, Zhenhua Yue, Jianke Li, Lele Sun, Astrid Irwanto, Yongxiang Yu, Mingfei Chen, Zihao Mi, Honglei Wang, Pengcheng Huai, Yi Li, Tiantian Du, Wenjun Yu, Yang Xia, Hailu Xiao, Jiabao You, Jinghui Li, Qing Yang, Na Wang, Panpan Shang, Guiye Niu, Xiaojun Chi, Xiuhuan Wang, Jing Cao, Xiujun Cheng, Hong Liu, Jianjun Liu, and Furen Zhang. A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nature Communications*, 7(1):1–8, 2016. ISSN 20411723. doi: 10.1038/ncomms13760.

[150] Xuelei Zhang, Yuyan Cheng, Qun Zhang, Xiaomeng Wang, Yan Lin, Chao Yang, Jingying Sun, He Huang, Yang Li, Yujun Sheng, Xing Fan, Yonghu Sun, Xuejun Zhang, Xiaodong Zheng, Bo Zhang, and Sen Yang. Meta-Analysis Identifies Major Histocompatiblity Complex Loci in or Near HLA-DRB1, HLA-DQA1, HLA-C as Associated with Leprosy in Chinese Han Population. *Journal of Investigative Dermatology*, 139(4):957–960, 2019. ISSN 15231747. doi: 10.1016/j.jid.2018.09.029.

[151] Jun Hirata, Kazuyoshi Hosomichi, Saori Sakaue, Masahiro Kanai, Hirofumi Nakaoka, Kazuyoshi Ishigaki, Ken Suzuki, Masato Akiyama, Toshihiro Kishikawa, Kotaro Ogawa, Tatsuo Masuda, Kenichi Yamamoto, Makoto Hirata, Koichi Matsuda, Yukihide Momozawa, Ituro Inoue, Michiaki Kubo, Yoichiro Kamatani, and Yukinori Okada. Genetic and phenotypic landscape of the major histocompatibilty complex region in the Japanese population. *Nature Genetics*, 51(3):470–480, 2019. ISSN 15461718. doi: 10.1038/s41588-018-0336-0.

[152] Ana Valencia, Candelaria Vergara, Chloe L. Thio, Nicolas Vince, Venceslas Douillard, Alba Grifoni, Andrea L. Cox, Eric O. Johnson, Alex H. Kral, James J. Goedert, Alessandra Mangia, Valeria Piazzolla, Shruti H. Mehta, Gregory D. Kirk, Arthur Y. Kim, Georg M. Lauer, Raymond T. Chung, Jennifer C. Price, Salim I. Khakoo, Laurent Alric, Matthew E. Cramp, Sharyne M. Donfield, Brian R. Edlin, Michael P. Busch, Graeme Alexander, Hugo R. Rosen, Edward L. Murphy, Genevieve L. Wojcik, Mary Carrington, Pierre Antoine Gourraud, Alessandro Sette, David L. Thomas, and Priya Duggal. Trans-ancestral fine-mapping of MHC reveals key amino acids associated with spontaneous clearance of hepatitis C in HLA-DQβ1. *American Journal of Human Genetics*, 109(2):299–310, 2022. ISSN 15376605. doi: 10.1016/j.ajhg.2022.01.001.

[153] Thomas Tängdén, Stefan Gustafsson, Abhiram S. Rao, and Erik Ingelsson. A genome-wide association study in a large community-based cohort identifies multiple loci associated with susceptibility to bacterial and viral infections. *Scientific Reports*, 12(1):1–14, 2022. ISSN 20452322. doi: 10.1038/s41598-022-05838-z.

[154] Paul J. McLaren, Stephan Ripke, Kimberly Pelak, Amy C. Weintrob, Nikolaos A. Patsopoulos, Xiaoming Jia, Rachel L. Erlich, Niall J. Lennon, Carl M. Kadie, David Heckerman, Namrata Gupta, David W. Haas, Steven G. Deeks, Florencia Pereyra, Bruce D. Walker, Paul I.W. de Bakker, Daniel R. Kuritzkes, Gregory K. Robbins, Robert W. Shafer, Roy M. Gulick, Cecilia M. Shikuma, Richard Haubrich, Sharon Riddler, Paul E. Sax, Eric S. Daar, Heather J. Ribaudo, Marylyn M. Addo, Brian Agan, Shanu Agarwal, Richard L. Ahern, Brady L. Allen, Sherly Altidor, Eric L. Altschuler, Sujata Ambardar, Kathryn Anastos, Val Anderson, Ushan Andrady, Diana Antoniskis, David Bangsberg, Daniel Barbaro, William Barrie, J. Bartczak, Simon Barton, Patricia Basden, Nesli Basgoz, Nicholaos C. Bellos, Judith Berger, Nicole F. Bernard, Annette M. Bernard, Stanley J. Bodner, Robert K. Bolan, Emilie T. Boudreaux, James F. Braun, Jon E. Brndjar, J. Brown, Sheldon T. Brown, Jedidiah Burack, Larry M. Bush, Virginia Cafaro, John Campbell, Robert H. Carlson, J. Kevin Carmichael, Kathleen K. Casey, Chris Cavacuiti, Gregory Celestin, Steven T. Chambers, Nancy Chez, Lisa M. Chirch, Paul J. Cimoch, Daniel Cohen, Lillian E. Cohn, Brian

Conway, David A. Cooper, Brian Cornelson, David T. Cox, Michael V. Cristofano, George Cuchural, Julie L. Czartoski, Joseph M. Dahman, Jennifer S. Daly, Benjamin T. Davis, Kristine Davis, Sheila M. Davod, Edwin Dejesus, Craig A. Dietz, Eleanor Dunham, Michael E. Dunn, Todd B. Ellerin, Joseph J. Eron, John J.W. Fangman, Helen Ferlazzo, Sarah Fidler, Anita Fleenor-Ford, Renee Frankel, Kenneth A. Freedberg, Neel K. French, Jonathan D. Fuchs, Jon D. Fuller, Jonna Gaberman, Joel E. Gallant, Rajesh T. Gandhi, Efrain Garcia, Donald Garmon, Joseph C. Gathe, Cyril R. Gaultier, Wondwoosen Gebre, Frank D. Gilman, Ian Gilson, Paul A. Goepfert, Michael S. Gottlieb, Claudia Goulston, Richard K. Groger, T. Douglas Gurley, Stuart Haber, Robin Hardwicke, W. David Hardy, P. Richard Harrigan, Trevor N. Hawkins, Sonya Heath, Frederick M. Hecht, W. Keith Henry, Melissa Hladek, Robert P. Hoffman, James M. Horton, Ricky K. Hsu, Gregory D. Huhn, Peter Hunt, Mark L. Illeman, Hans Jaeger, Robert M. Jellinger, Mina John, Jennifer A. Johnson, Kristin L. Johnson, Heather Johnson, Kay Johnson, Jennifer Joly, Wilbert C. Jordan, Carol A. Kauffman, Homayoon Khanlou, Arthur Y. Kim, David D. Kim, Clifford A. Kinder, Laura Kogelman, Erna Milunka Kojic, P. Todd Korthuis, Wayne Kurisu, Douglas S. Kwon, Melissa Lamar, Harry Lampiris, Michael M. Lederman, David M. Lee, Marah J. Lee, Edward T.Y. Lee, Janice Lemoine, Jay A. Levy, Josep M. Llibre, Michael A. Liguori, Susan J. Little, Anne Y. Liu, Alvaro J. Lopez, Mono R. Loutfy, Dawn Loy, Debbie Y. Mohammed, Alan Man, Michael K. Mansour, Vincent C. Marconi, Martin Markowitz, Jeffrey N. Martin, Harold L. Martin, Kenneth Hugh Mayer, M. Juliana McElrath, Theresa A. McGhee, Barbara H. McGovern, Katherine McGowan, Dawn McIntyre, Gavin X. McLeod, Prema Menezes, Greg Mesa, Craig E. Metroka, Dirk Meyer-Olson, Andy O. Miller, Kate Montgomery, Karam C. Mounzer, Iris Nagin, Ronald G. Nahass, Craig Nielsen, David L. Norene, David H. O'connor, Jason Okulicz, Edward C. Oldfield, Susan A. Olender, Mario Ostrowski, William F. Owen, Jeffrey Parsonnet, Andrew M. Pavlatos, Alicja Piechocka-Trocha, Aaron M. Perlmutter, Jonathan M. Pincus, Leandro Pisani, Lawrence Jay Price, Laurie Proia, Richard C. Prokesch, Heather Calderon Pujet, Moti Ramgopal, Michael Rausch, J. Ravishankar, Frank S. Rhame, Constance Shamuyarira Richards, Douglas D. Richman, Berta Rodes, Milagros Rodriguez, Richard C. Rose, Eric S. Rosenberg, Daniel Rosenthal, Polly E. Ross, David S. Rubin, Elease Rumbaugh, Luis Saenz, Michelle R. Salvaggio, William C. Sanchez, Veeraf M. Sanjana, Steven Santiago, Wolfgang Schmidt, Hanneke Schuitemaker, Philip M. Sestak, Peter Shalit, William Shay, Vivian N. Shirvani, Vanessa I. Silebi, James M. Sizemore, Paul R. Skolnik, Marcia Sokol-Anderson, James M. Sosman, Paul Stabile, Jack T. Stapleton, Francine Stein, Hans Jurgen Stellbrink, F. Lisa Sterman, Valerie E. Stone, David R. Stone, Giuseppe Tambussi, Randy A. Taplitz, Ellen M. Tedaldi, Amalio Telenti, Richard Torres, Lorraine Tosiello, Cecile Tremblay, Marc A. Tribble, Phuong D. Trinh, Anthony Vaccaro, Emilia Valadas, Thanes J. Vanig, Isabel Vecino, Wenoah Veikley, Barbara H. Wade, Charles Walworth, Chingchai Wanidworanun, Douglas J. Ward, Robert D. Weber, Duncan Webster, Steve Weis, David A. Wheeler, David J. White, Ed Wilkins, Alan Winston, Clifford G. Wlodaver, Angelique Van'T Wout, David P. Wright, Otto O. Yang, David L. Yurdin, Brandon W. Zabukovic, Kimon C. Zachary, Beth Zeeman, and Meng Zhao. Fine-mapping classical HLA variation associated with durable host control of HIV-1 infection in african americans. *Human Molecular Genetics*, 21(19):4334–4347, 2012. ISSN 14602083. doi: 10.1093/hmg/dds226.

[155] Regina C. Betz, Lynn Petukhova, Stephan Ripke, Hailiang Huang, Androniki Menelaou, Silke Redler, Tim Becker, Stefanie Heilmann, Tarek Yamany, Madeliene Duvic, Maria Hordinsky, David Norris, Vera H. Price, Julian MacKay-Wiggan, Annemieke De Jong, Gina M. DeStefano, Susanne Moebus, Markus Böhm, Ulrike Blume-Peytavi, Hans Wolff, Gerhard Lutz, Roland Kruse, Li Bian, Christopher I. Amos, Annette Lee, Peter K. Gregersen, Bettina Blaumeiser, David Altshuler, Raphael Clynes, Paul I.W. De Bakker, Markus M. Nöthen, Mark J. Daly, and Angela M. Christiano. Genome-wide meta-analysis in alopecia areata resolves HLA associations and reveals two new susceptibility loci. *Nature Communications*, 6(1):1–8, 2015. ISSN 20411723. doi: 10.1038/ncomms6966.

[156] Yukinori Okada, Yukihide Momozawa, Kyota Ashikawa, Masahiro Kanai, Koichi Matsuda, Yoichiro Kamatani, Atsushi Takahashi, and Michiaki Kubo. Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nature Genetics*, 47(7): 798–802, 2015. ISSN 15461718. doi: 10.1038/ng.3310. URL `http://dx.doi.org/10.1038/ng.3310`.

[157] Xinli Hu, Aaron J. Deutsch, Tobias L. Lenz, Suna Onengut-Gumuscu, Buhm Han, Wei Min Chen, Joanna M.M. Howson, John A. Todd, Paul I.W. De Bakker, Stephen S. Rich, and Soumya Raychaudhuri. Additive and interaction effects at three amino acid positions in HLA-DQ and HLA-DR molecules drive type 1 diabetes risk. *Nature Genetics*, 47(8):898–905, 2015. ISSN 15461718. doi: 10.1038/ng.3353.

[158] Tatsuhiko Naito, Ken Suzuki, Jun Hirata, Yoichiro Kamatani, Koichi Matsuda, Tatsushi Toda, and Yukinori Okada. A deep learning method for HLA imputation and trans-ethnic MHC fine-mapping of type 1 diabetes. *Nature Communications*, 12(1):1–14, 2021. ISSN 20411723. doi: 10.1038/s41467-021-21975-x.

[159] Lu Cao, Ruixue Zhang, Yirui Wang, Xia Hu, Liang Yong, Bao Li, Huiyao Ge, Weiwei Chen, Qi Zhen, Yafen Yu, Yiwen Mao, Zhuo Li, Wencheng Fan, and Liangdan Sun. Fine Mapping Analysis of the MHC Region to Identify Variants Associated With Chinese Vitiligo and SLE and Association Across These Diseases. *Frontiers in Immunology*, 12:1–7, 2022. ISSN 16643224. doi: 10.3389/fimmu.2021.758652.

[160] Hanna M. Ollila, Jean Marie Ravel, Fang Han, Juliette Faraco, Ling Lin, Xiuwen Zheng, Giuseppe Plazzi, Yves Dauvilliers, Fabio Pizza, Seung Chul Hong, Poul Jennum, Stine Knudsen, Birgitte R. Kornum, Xiao Song Dong, Han Yan, Heeseung Hong, Cristin Coquillard, Joshua Mahlios, Otto Jolanki, Mali Einen, Sophie Lavault, Birgit Högl, Birgit Frauscher, Catherine Crowe, Markku Partinen, Yu Shu Huang, Patrice Bourgin, Outi Vaarala, Alex Désautels, Jacques Montplaisir, Steven J. Mack, Michael Mindrinos, Marcelo Fernandez-Vina, and Emmanuel Mignot. HLA-DPB1 and HLA class i confer risk of and protection from narcolepsy. *American Journal of Human Genetics*, 96(1): 136–146, 2015. ISSN 15376605. doi: 10.1016/j.ajhg.2014.12.010.

[161] Chan Wang, Xiaodong Zheng, Ruqi Tang, Chongxu Han, Yuzhang Jiang, Jian Wu, Youlin Shao, Yueqiu Gao, Jianjiang Yu, Zhigang Hu, Zhidong Zang, Yi Zhao, Na Dai, Lei Liu, Xudong Wu, Jinshan Nie, Bo Jiang, Maosong Lin, Li Li, Yiran Wei, You Li, Yuhua Gong, Yaping Dai, Lan Wang, Ningling Ding, Ping Xu, Sufang Chen, Peng Jiang, Lu Wang, Fang Qiu, Qiuyuan Wu, Mingming Zhang, Rohil Jawed, Ru Chen, Yu Zhang, Xingjuan Shi, Zhen Zhu, Hao Pei, Lihua Huang, Ye Tian, Kui Zhang, Hong Qiu, Weifeng Zhao, M. Eric Gershwin, Weichang Chen, Michael F. Seldin, Xiangdong Liu, Xiong Ma, and Liangdan Sun. Fine mapping of the MHC region identifies major independent variants associated with Han Chinese primary biliary cholangitis. *Journal of Autoimmunity*, 107:102372, 2020. ISSN 10959157. doi: 10.1016/j.jaut.2019.102372.

[162] Michio Yasunami, Hitomi Nakamura, Katsushi Tokunaga, Minae Kawashima, Nao Nishida, Yuki Hitomi, and Minoru Nakamura. Principal contribution of HLA-DQ alleles, DQB106:04 and DQB103:01, to disease resistance against primary biliary cholangitis in a Japanese population. *Scientific Reports*, 7(1):1–10, 2017. ISSN 20452322. doi: 10.1038/s41598-017-11148-6.

[163] Rebecca Darlay, Kristin L. Ayers, George F. Mells, Lynsey S. Hall, Jimmy Z. Liu, Mohamed A. Almarri, Graeme J. Alexander, David E. Jones, Richard N. Sandford, Carl A. Anderson, and Heather J. Cordell. Amino acid residues in five separate HLA genes can explain most of the known associations between the MHC and primary biliary cholangitis. *PLoS Genetics*, 14(12):1–29, 2018. ISSN 15537404. doi: 10.1371/journal.pgen.1007833.

[164] Javier Gutierrez-Achury, Alexandra Zhernakova, Sara L. Pulit, Gosia Trynka, Karen A. Hunt, Jihane Romanos, Soumya Raychaudhuri, David A. Van Heel, Cisca Wijmenga, and Paul I.W. De Bakker. Fine mapping in the MHC region accounts for 18% additional genetic risk for celiac disease. *Nature Genetics*, 47(6):577–578, 2015. ISSN 15461718. doi: 10.1038/ng.3268.

[165] Philippe Goyette, Gabrielle Boucher, Dermot Mallon, Eva Ellinghaus, Luke Jostins, Hailiang Huang, Stephan Ripke, Elena S. Gusareva, Vito Annese, Stephen L. Hauser, Jorge R. Oksenberg, Ingo Thomsen, Stephen Leslie, Mark J. Daly, Kristel Van Steen, Richard H. Duerr, Jeffrey C. Barrett, Dermot P.B. McGovern, L. Philip Schumm, James A. Traherne, Mary N. Carrington, Vasilis Kosmoliaptsis, Tom H. Karlsen, Andre Franke, John D. Rioux, Clara Abraham, Jean Paul Achkar,

Tariq Ahmad, Leila Amininejad, Ashwin N. Ananthakrishnan, Vibeke Andersen, Carl A. Anderson, Jane M. Andrews, Guy Aumais, Leonard Baidoo, Robert N. Baldassano, Tobias Balschun, Peter A. Bampton, Murray Barclay, Theodore M. Bayless, Johannes Bethge, Joshua C. Bis, Alain Bitton, Stephan Brand, Steven R. Brant, Carsten Buning, Angela Chew, Judy H. Cho, Isabelle Cleynen, Ariella Cohain, Anthony Croft, Mauro D'Amato, Silvio Danese, Dirk De Jong, Martine De Vos, Goda Denapiene, Lee A. Denson, Kathy L. Devaney, Olivier Dewit, Renata D'Inca, Marla Dubinsky, Cathryn Edwards, David Ellinghaus, Jonah Essers, Lynnette R. Ferguson, Eleonora A. Festen, Philip Fleshner, Tim Florin, Denis Franchimont, Karin Fransen, Richard Gearry, Michel Georges, Christian Gieger, Jurgen Glas, Todd Green, Anne M. Griffiths, Stephen L. Guthery, Hakon Hakonarson, Jonas Halfvarson, Katherine Hanigan, Talin Haritunians, Ailsa Hart, Chris Hawkey, Nicholas K. Hayward, Matija Hedl, Paul Henderson, Xinli Hu, Ken Y. Hui, Marcin Imielinski, Andrew Ippoliti, Laimas Jonaitis, Nicholas A. Kennedy, Mohammed Azam Khan, Gediminas Kiudelis, Subra Kugathasan, Limas Kupcinskas, Anna Latiano, Debby Laukens, Ian C. Lawrance, James C. Lee, Charlie W. Lees, Marcis Leja, Johan Van Limbergen, Paolo Lionetti, Jimmy Z. Liu, Edouard Louis, Gillian Mahy, John Mansfield, Dunecan Massey, Christopher G. Mathew, Raquel Milgrom, Mitja Mitrovic, Grant Montgomery, Craig Mowat, Wwilliam Newman, Aylwin Ng, Siew C. Ng, Sok Meng Evelyn Ng, Susanna Nikolaus, Kaida Ning, Markus Nothen, Ioannis Oikonomou, Orazio Palmieri, Miles Parkes, Anne Phillips, Cyriel Y. Ponsioen, Uros Potocnik, Natalie J. Prescott, Deborah D. Proctor, Graham Radford-Smith, Jean Francois Rahier, Soumya Raychaudhur, Miguel Regueiro, Florian Rieder, Rebecca Roberts, Richard K. Russell, Jeremy D. Sanderson, Miquel Sans, Jack Satsangi, Eric E. Schadt, Stefan Schreiber, L. Philip Schumm, Regan Scott, Mark Seielstad, Yashoda Sharma, Mark S. Silverberg, Lisa A. Simms, Jurgita Skieceviciene, Sarah L. Spain, A. Hillary Steinhart, Joanne M. Stempak, Laura Stronati, Jurgita Sventoraityte, Stephan R. Targan, Kirstin M. Taylor, Anje Ter Velde, Emilie Theatre, Leif Torkvist, Mark Tremelling, Andrea Van Der Meulen, Suzanne Van Sommeren, Eric Vasiliauskas, Severine Vermeire, Hein W. Verspaget, Thomas Walters, Kai Wwang, Ming Hsi Wwang, Rinse K. Wweersma, Zhi Wei, David Whiteman, Cisca Wijmenga, David C. Wilson, Juliane Winkelmann, Ramnik J. Xavier, Sebastian Zeissig, Bin Zhang, Clarence K. Zhang, Hu Zhang, Wwei Zhang, Hongyu Zhao, and Zhen Z. Zhao. High-density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nature Genetics*, 47(2):172–179, 2015. ISSN 15461718. doi: 10.1038/ng.3176.

[166] A. Hinks, J. Bowes, J. Cobb, H. C. Ainsworth, M. C. Marion, M. E. Comeau, M. Sudman, B. Han, M. L. Becker, J. F. Bohnsack, P. I.W. De Bakker, J. P. Haas, M. Hazen, D. J. Lovell, P. A. Nigrovic, E. Nordal, M. Punnaro, A. M. Rosenberg, M. Rygg, S. L. Smith, C. A. Wise, V. Videm, L. R. Wedderburn, A. Yarwood, R. S.M. Yeung, S. Prahalad, C. D. Langefeld, S. Raychaudhuri, S. D. Thompson, and W. Thomson. Fine-mapping the MHC locus in juvenile idiopathic arthritis (JIA) reveals genetic heterogeneity corresponding to distinct adult inflammatory arthritic diseases. *Annals of the Rheumatic Diseases*, 76(4):765–772, 2017. ISSN 14682060. doi: 10.1136/annrheumdis-2016-210025.

[167] Soumya Raychaudhuri, Cynthia Sandor, Eli A. Stahl, Jan Freudenberg, Hye Soon Lee, Xiaoming Jia, Lars Alfredsson, Leonid Padyukov, Lars Klareskog, Jane Worthington, Katherine A. Siminovitch, Sang Cheol Bae, Robert M. Plenge, Peter K. Gregersen, and Paul I.W. De Bakker. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genetics*, 44(3):291–296, 2012. ISSN 10614036. doi: 10.1038/ng.1076.

[168] Yukinori Okada, Buhm Han, Lam C. Tsoi, Philip E. Stuart, Eva Ellinghaus, Trilokraj Tejasvi, Vinod Chandran, Fawnda Pellett, Remy Pollock, Anne M. Bowcock, Gerald G. Krueger, Michael Weichenthal, John J. Voorhees, Proton Rahman, Peter K. Gregersen, Andre Franke, Rajan P. Nair, Gonçalo R. Abecasis, Dafna D. Gladman, James T. Elder, Paul I.W. De Bakker, and Soumya Raychaudhuri. Fine mapping major histocompatibility complex associations in psoriasis and its clinical subtypes. *American Journal of Human Genetics*, 95(2):162–172, 2014. ISSN 15376605. doi: 10.1016/j.ajhg.2014.07.002.

[169] Fusheng Zhou, Hongzhi Cao, Xianbo Zuo, Tao Zhang, Xiaoguang Zhang, Xiaomin Liu, Ricong Xu, Gang Chen, Yuanwei Zhang, Xiaodong Zheng, Xin Jin, Jinping Gao, Junpu Mei, Yujun Sheng,

Qibin Li, Bo Liang, Juan Shen, Changbing Shen, Hui Jiang, Caihong Zhu, Xing Fan, Fengping Xu, Min Yue, Xianyong Yin, Chen Ye, Cuicui Zhang, Xiao Liu, Liang Yu, Jinghua Wu, Mengyun Chen, Xuehan Zhuang, Lili Tang, Haojing Shao, Longmao Wu, Jian Li, Yu Xu, Yijie Zhang, Suli Zhao, Yu Wang, Ge Li, Hanshi Xu, Lei Zeng, Jianan Wang, Mingzhou Bai, Yanling Chen, Wei Chen, Tian Kang, Yanyan Wu, Xun Xu, Zhengwei Zhu, Yong Cui, Zaixing Wang, Chunjun Yang, Peiguang Wang, Leihong Xiang, Xiang Chen, Anping Zhang, Xinghua Gao, Furen Zhang, Jinhua Xu, Min Zheng, Jie Zheng, Jianzhong Zhang, Xueqing Yu, Yingrui Li, Sen Yang, Huanming Yang, Jian Wang, Jianjun Liu, Lennart Hammarström, Liangdan Sun, Jun Wang, and Xuejun Zhang. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nature Genetics*, 48(7):740–746, 2016. ISSN 15461718. doi: 10.1038/ng.3576.

[170] Philip E. Stuart, Lam C. Tsoi, Rajan P. Nair, Manju Ghosh, Madhulika Kabra, Pakeeza A. Shaiq, Ghazala K. Raja, Raheel Qamar, B. K. Thelma, Matthew T. Patrick, Anita Parihar, Sonam Singh, Sujay Khandpur, Uma Kumar, Michael Wittig, Frauke Degenhardt, Trilokraj Tejasvi, John J. Voorhees, Stephan Weidinger, Andre Franke, Goncalo R. Abecasis, Vinod K. Sharma, and James T. Elder. Transethnic analysis of psoriasis susceptibility in South Asians and Europeans enhances fine mapping in the MHC and genome wide. *Human Genetics and Genomics Advances*, 3(1):100069, 2022. ISSN 26662477. doi: 10.1016/j.xhgg.2021.100069.

[171] Marialbert Acosta-Herrera, Martin Kerick, Elena Lopéz-Isac, Shervin Assassi, Lorenzo Beretta, Carmen Pilar Simeón-Aznar, Norberto Ortego-Centeno, Susanna M. Proudman, Nicolas Hunzelmann, Gianluca Moroncini, Jeska K. De Vries-Bouwstra, Gisela Orozco, Anne Barton, Ariane L. Herrick, Chikashi Terao, Yannick Allanore, Matthew A. Brown, Timothy R.D.J. Radstake, Carmen Fonseca, Christopher P. Denton, Maureen D. Mayes, and Javier Martin. Comprehensive analysis of the major histocompatibility complex in systemic sclerosis identifies differential HLA associations by clinical and serological subtypes. *Annals of the Rheumatic Diseases*, 80(8):1040–1047, 2021. ISSN 14682060. doi: 10.1136/annrheumdis-2021-219884.

[172] Nikolaos A. Patsopoulos, Sergio E. Baranzini, Adam Santaniello, Parisa Shoostari, Chris Cotsapas, Garrett Wong, Ashley H. Beecham, Tojo James, Joseph Replogle, Ioannis S. Vlachos, Cristin McCabe, Tune H. Pers, Aaron Brandes, Charles White, Brendan Keenan, Maria Cimpean, Phoebe Winn, Ioannis Pavlos Panteliadis, Allison Robbins, Till F.M. Andlauer, Onigiusz Zarzycki, Bénédicte Dubois, An Goris, Helle Bach Søndergaard, Finn Sellebjerg, Per Soelberg Sorensen, Henrik Ullum, Lise Wegner Thørner, Janna Saarela, Isabelle Cournu-Rebeix, Vincent Damotte, Bertrand Fontaine, Lena Guillot-Noel, Mark Lathrop, Sandra Vukusic, Achim Berthele, Viola Pongratz, Dorothea Buck, Christiane Gasperi, Christiane Graetz, Verena Grummel, Bernhard Hemmer, Muni Hoshi, Benjamin Knier, Thomas Korn, Christina M. Lill, Felix Luessi, Mark Mühlau, Frauke Zipp, Efthimios Dardiotis, Cristina Agliardi, Antonio Amoroso, Nadia Barizzone, Maria D. Benedetti, Luisa Bernardinelli, Paola Cavalla, Ferdinando Clarelli, Giancarlo Comi, Daniele Cusi, Federica Esposito, Laura Ferrè, Daniela Galimberti, Clara Guaschino, Maurizio A. Leone, Vittorio Martinelli, Lucia Moiola, Marco Salvetti, Melissa Sorosina, Domizia Vecchio, Andrea Zauli, Silvia Santoro, Nicasio Mancini, Miriam Zuccalà, Julia Mescheriakova, Cornelia Van Duijn, Steffan D. Bos, Elisabeth G. Celius, Anne Spurkland, Manuel Comabella, Xavier Montalban, Lars Alfredsson, Izaura L. Bomfim, David Gomez-Cabrero, Jan Hillert, Maja Jagodic, Magdalena Lindén, Fredrik Piehl, Ilijas Jelčić, Roland Martin, Mirela Sospedra, Amie Baker, Maria Ban, Clive Hawkins, Pirro Hysi, Seema Kalra, Fredrik Karpe, Jyoti Khadake, Genevieve Lachance, Paul Molyneux, Matthew Neville, John Thorpe, Elizabeth Bradshaw, Stacy J. Caillier, Peter Calabresi, Bruce A.C. Cree, Anne Cross, Mary Davis, Paul W.I. De Bakker, Silvia Delgado, Marieme Dembele, Keith Edwards, Kate Fitzgerald, Irene Y. Frohlich, Pierre Antoine Gourraud, Jonathan L. Haines, Hakon Hakonarson, Dorlan Kimbrough, Noriko Isobe, Ioanna Konidari, Ellen Lathi, Michelle H. Lee, Taibo Li, David An, Andrew Zimmer, Lohith Madireddy, Clara P. Manrique, Mitja Mitrovic, Marta Olah, Ellis Patrick, Margaret A. Pericak-Vance, Laura Piccio, Cathy Schaefer, Howard Weiner, Kasper Lage, Alastair Compston, David Hafler, Hanne F. Harbo, Stephen L. Hauser, Graeme Stewart, Sandra D'Alfonso, Georgios Hadjigeorgiou, Bruce Taylor, Lisa F. Barcellos, David Booth, Rogier Hintzen, Ingrid Kockum, Filippo Martinelli-Boneschi, Jacob L. McCauley, Jorge R. Oksenberg, Annette Oturai, Stephen Sawcer,

Adrian J. Ivinson, Tomas Olsson, and Philip L. De Jager. Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. *Science*, 365(6460), 2019. ISSN 10959203. doi: 10.1126/science.aav7188.

[173] Julio E. Molineros, Loren L. Looger, Kwangwoo Kim, Yukinori Okada, Chikashi Terao, Celi Sun, Xu jie Zhou, Prithvi Raj, Yuta Kochi, Akari Suzuki, Shuji Akizuki, Shuichiro Nakabo, So Young Bang, Hye Soon Lee, Young Mo Kang, Chang Hee Suh, Won Tae Chung, Yong Beom Park, Jung Yoon Choe, Seung Cheol Shim, Shin Seok Lee, Xiaoxia Zuo, Kazuhiko Yamamoto, Quan Zhen Li, Nan Shen, Lauren L. Porter, John B. Harley, Kek Heng Chua, Hong Zhang, Edward K. Wakeland, Betty P. Tsao, Sang Cheol Bae, and Swapan K. Nath. Amino acid signatures of HLA Class-I and II molecules are strongly associated with SLE susceptibility and autoantibody production in Eastern Asians. *PLoS Genetics*, 15(4):1–25, 2019. ISSN 15537404. doi: 10.1371/journal.pgen.1008092.

[174] Johannes Waage, Marie Standl, John A. Curtin, Leon E. Jessen, Jonathan Thorsen, Chao Tian, Nathan Schoettler, Carlos Flores, Abdel Abdellaoui, Tarunveer S. Ahluwalia, Alexessander C. Alves, Andre F.S. Amaral, Josep M. Antó, Andreas Arnold, Amalia Barreto-Luis, Hansjörg Baurecht, Catharina E.M. van Beijsterveldt, Eugene R. Bleecker, Sílvia Bonàs-Guarch, Dorret I. Boomsma, Susanne Brix, Supinda Bunyavanich, Esteban G. Burchard, Zhanghua Chen, Ivan Curjuric, Adnan Custovic, Herman T. den Dekker, Shyamali C. Dharmage, Julia Dmitrieva, Liesbeth Duijts, Markus J. Ege, W. James Gauderman, Michel Georges, Christian Gieger, Frank Gilliland, Raquel Granell, Hongsheng Gui, Torben Hansen, Joachim Heinrich, John Henderson, Natalia Hernandez-Pacheco, Patrick Holt, Medea Imboden, Vincent W.V. Jaddoe, Marjo Riitta Jarvelin, Deborah L. Jarvis, Kamilla K. Jensen, Ingileif Jónsdóttir, Michael Kabesch, Jaakko Kaprio, Ashish Kumar, Young Ae Lee, Albert M. Levin, Xingnan Li, Fabian Lorenzo-Diaz, Erik Melén, Josep M. Mercader, Deborah A. Meyers, Rachel Myers, Dan L. Nicolae, Ellen A. Nohr, Teemu Palviainen, Lavinia Paternoster, Craig E. Pennell, Göran Pershagen, Maria Pino-Yanes, Nicole M. Probst-Hensch, Franz Rüschendorf, Angela Simpson, Kari Stefansson, Jordi Sunyer, Gardar Sveinbjornsson, Elisabeth Thiering, Philip J. Thompson, Maties Torrent, David Torrents, Joyce Y. Tung, Carol A. Wang, Stephan Weidinger, Scott Weiss, Gonneke Willemsen, L. Keoki Williams, Carole Ober, David A. Hinds, Manuel A. Ferreira, Hans Bisgaard, David P. Strachan, and Klaus Bønnelykke. Genome-wide association and HLA fine-mapping studies identify risk loci and genetic pathways underlying allergic rhinitis. *Nature Genetics*, 50(8):1072–1080, 2018. ISSN 15461718. doi: 10.1038/s41588-018-0157-1.

[175] Kristi Krebs, Jonas Bovijn, Neil Zheng, Maarja Lepamets, Jenny C. Censin, Tuuli Jürgenson, Dage Särg, Erik Abner, Triin Laisk, Yang Luo, Line Skotte, Frank Geller, Bjarke Feenstra, Wei Wang, Adam Auton, Michelle Agee, Stella Aslibekyan, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sarah L. Elson, Kipper Fletez-Brant, Pierre Fontanillas, Nicholas A. Furlotte, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng Han Lin, Nadia K. Litterman, Marie K. Luff, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Aaron A. Petrakovitz, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Anjali J. Shastri, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, Xin Wang, Amir S. Zare, Soumya Raychaudhuri, Tõnu Esko, Andres Metspalu, Sven Laur, Dan M. Roden, Wei Qi Wei, Michael V. Holmes, Cecilia M. Lindgren, Elizabeth J. Phillips, Reedik Mägi, Lili Milani, and João Fadista. Genome-wide Study Identifies Association between HLA-B55:01 and Self-Reported Penicillin Allergy. *American Journal of Human Genetics*, 107(4):612–621, 2020. ISSN 15376605. doi: 10.1016/j.ajhg.2020.08.008.

[176] Aida Ferreiro-Iglesias, Corina Lesseur, James McKay, Rayjean J. Hung, Younghun Han, Xuchen Zong, David Christiani, Mattias Johansson, Xiangjun Xiao, Yafang Li, David C. Qian, Xuemei Ji, Geoffrey Liu, Neil Caporaso, Ghislaine Scelo, David Zaridze, Anush Mukeriya, Milica Kontic, Simona Ognjanovic, Jolanta Lissowska, Małgorzata Szołkowska, Beata Swiatkowska, Vladimir Janout, Ivana Holcatova, Ciprian Bolca, Milan Savic, Miodrag Ognjanovic, Stig Egil Bojesen, Xifeng Wu, Demetrios Albanes, Melinda C. Aldrich, Adonina Tardon, Ana Fernandez-Somoano,

Guillermo Fernandez-Tardon, Loic Le Marchand, Gadi Rennert, Chu Chen, Jennifer Doherty, Gary Goodman, Heike Bickeböller, H. Erich Wichmann, Angela Risch, Albert Rosenberger, Hongbing Shen, Juncheng Dai, John K. Field, Michael Davies, Penella Woll, M. Dawn Teare, Lambertus A. Kiemeney, Erik H.F.M. van der Heijden, Jian Min Yuan, Yun Chul Hong, Aage Haugen, Shanbeh Zienolddiny, Stephen Lam, Ming Sound Tsao, Mikael Johansson, Kjell Grankvist, Matthew B. Schabath, Angeline Andrew, Eric Duell, Olle Melander, Hans Brunnström, Philip Lazarus, Susanne Arnold, Stacey Slone, Jinyoung Byun, Ahsan Kamal, Dakai Zhu, Maria Teresa Landi, Christopher I. Amos, and Paul Brennan. Fine mapping of MHC region in lung cancer highlights independent susceptibility loci by ethnicity. *Nature Communications*, 9(1):1–12, 2018. ISSN 20411723. doi: 10.1038/s41467-018-05890-2.

[177] Kazuyoshi Ishigaki, Kaitlyn A. Lagattuta, Yang Luo, Eddie A. James, Jane H. Buckner, and Soumya Raychaudhuri. HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors. *Nature Genetics*, 54(4):393–402, 2022. ISSN 15461718. doi: 10.1038/s41588-022-01032-z.

[178] Eilon Sharon, Leah V. Sibener, Alexis Battle, Hunter B. Fraser, K. Christopher Garcia, and Jonathan K. Pritchard. Genetic variation in MHC proteins is associated with T cell receptor expression biases. *Nature Genetics*, 48(9):995–1002, 2016. ISSN 15461718. doi: 10.1038/ng.3625.

[179] M. A. Saper, P. J. Bjorkman, and D. C. Wiley. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *Journal of Molecular Biology*, 219(2):277–319, 1991. ISSN 00222836. doi: 10.1016/0022-2836(91)90567-P.

[180] Masazumi Matsumura, Daved H. Fremont, Per A. Peterson, and Ian A. Wilson. Emerging principles for the recognition of peptide antigens by MHC class I molecules. *Science*, 257(5072):927–934, 1992. ISSN 00368075. doi: 10.1126/science.1323878.

[181] Gareth Chelvanayagam. A roadmap for HLA-A, HLA-B, and HLA-C peptide binding specificities. *Immunogenetics*, 45(1):15 – 26, 1996. doi: 10.1007/s002510050162.

[182] Xiaojiang Gao, Sue Lester, Anthony Veale, Barry Boettcher, Bart Currie, James McCluskey, and Gareth Chelvanayagam. HLA class I alleles in Australian aborigines and their peptide binding profiles. In Masanori Kasahara, editor, *Major Histocompatibility Complex: Evolution, Structure, and Function*, pages 446–462, Tokyo, 2000. Springer Japan. doi: 10.1007/978-4-431-65868-9.

[183] Gema Díaz, Massimo Amicosante, Dolores Jaraquemada, Richard H. Butler, M. Victoria Guillén, Miguel Sánchez, César Nombela, and Javier Arroyo. Functional analysis of HLA-DP polymorphism: A crucial role for DPβ residues 9, 11, 35, 55, 56, 69 and 84-87 in T cell allorecognition and peptide binding. *International Immunology*, 15(5):565–576, 2003. ISSN 09538178. doi: 10.1093/intimm/dxg057.

[184] Shaodong Dai, Guinevere A. Murphy, Frances Crawford, Douglas G. Mack, Michael T. Falta, Philippa Marrack, John W. Kappler, and Andrew P. Fontenot. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16):7425–7430, 2010. ISSN 00278424. doi: 10.1073/pnas.1001772107.

[185] J. Routsias and G. K. Papadopoulos. Polymorphic structural features of modelled HLA-DQ molecules segregate according to susceptibility or resistance to IDDM. *Diabetologia*, 38(11):1251–1261, 1995. ISSN 0012186X. doi: 10.1007/BF00401756.

[186] E. Yvonne Jones, Lars Fugger, Jack L. Strominger, and Christian Siebold. MHC class II proteins and disease: A structural perspective. *Nature Reviews Immunology*, 6(4):271–282, 2006. ISSN 14741733. doi: 10.1038/nri1805.

[187] George P. Bondinas, Antonis K. Moustakas, and George K. Papadopoulos. The spectrum of HLA-DQ and HLA-DR alleles, 2006: A listing correlating sequence and structure with function. *Immunogenetics*, 59(7):539–553, 2007. ISSN 00937711. doi: 10.1007/s00251-007-0224-8.

[188] Lawrence J. Stern, Jerry H. Brown, Theodore S. Jardetzky, Joan C. Gorga, Robert G. Urban, Jack L. Strominger, and Don C. Wiley. Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature*, 368(6468):215–221, 1994. ISSN 00280836. doi: 10.1038/368215a0.

[189] Ivan Dimitrov, Panayot Garnev, Darren R. Flower, and Irini Doytchinova. Peptide binding to the HLA-DRB1 supertype: A proteochemometrics analysis. *European Journal of Medicinal Chemistry*, 45(1):236–243, 2010. ISSN 02235234. doi: 10.1016/j.ejmech.2009.09.049.

[190] James Boocock, Meru J. Sadhu, Arun Durvasula, Joshua S. Bloom, and Leonid Kruglyak. Ancient balancing selection maintains incompatible versions of the galactose pathway in yeast. *Science*, 371 (6527):415–419, 2021. ISSN 10959203. doi: 10.1126/science.aba0542.

[191] Chris Todd Hittinger, Paula Gonçalves, José Paulo Sampaio, Jim Dover, Mark Johnston, and Antonis Rokas. Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature*, 464 (7285):54–58, 2010. ISSN 14764687. doi: 10.1038/nature08791.

[192] Linda T.A. Van Diepen, Ake Olson, Katarina Ihrmark, Jan Stenlid, and Timothy Y. James. Extensive trans-specific polymorphism at the mating type locus of the root decay fungus heterobasidion. *Molecular Biology and Evolution*, 30(10):2286–2301, 2013. ISSN 07374038. doi: 10.1093/molbev/mst126.

[193] Jennifer Wu, Sven J. Saupe, and N. Louise Glass. Evidence for balancing selection operating at the het-c heterokaryon incompatibility locus in a group of filamentous fungi. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21):12398–12403, 1998. ISSN 00278424. doi: 10.1073/pnas.95.21.12398.

[194] Benjamin Devier, Gabriela Aguileta, Michael E. Hood, and Tatiana Giraud. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics*, 181(1):209–223, 2009. ISSN 00166731. doi: 10.1534/genetics.108.093708.

[195] Talia L. Karasov, Joel M. Kniskern, Liping Gao, Brody J. Deyoung, Jing Ding, Ullrich Dubiella, Ruben O. Lastra, Sumitha Nallu, Fabrice Roux, Roger W. Innes, Luke G. Barrett, Richard R. Hudson, and Joy Bergelson. The long-term maintenance of a resistance polymorphism through diffuse interactions. *Nature*, 512(7515):436–440, 2014. ISSN 14764687. doi: 10.1038/nature13439.

[196] Eli A Stahl, Greg Dwyer, Rodney Mauricio, Martin Kreitman, and Joy Bergelson. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. *Nature*, 400(6745):667–671, 1999.

[197] Helen J. Brabham, Inmaculada Hernández-Pinzón, Samuel Holden, Jennifer Lorang, and Matthew J. Moscou. An ancient integration in a plant NLR is maintained as a trans-species polymorphism. *bioRxiv*, 2018. doi: 10.1101/239541.

[198] Qiong Wu, Ting Shen Han, Xi Chen, Jia Fu Chen, Yu Pan Zou, Zi Wen Li, Yong Chao Xu, and Ya Long Guo. Long-term balancing selection contributes to adaptation in Arabidopsis and its relatives. *Genome Biology*, 18(1):1–15, 2017. ISSN 1474760X. doi: 10.1186/s13059-017-1342-8.

[199] Marie Bouillé and Jean Bousquet. Trans-species shared polymorphisms at orthologous nuclear gene loci among distant species in the conifer Picea (Pinaceae): Implications for the long-term maintenance of genetic diversity in trees. *American Journal of Botany*, 92(1):63–73, 2005. ISSN 00029122. doi: 10.3732/ajb.92.1.63.

[200] Soochin Cho, Zachary Y. Huang, Daniel R. Green, Deborah R. Smith, and Jianzhi Zhang. Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research*, 16(11):1366–1375, 2006. ISSN 10889051. doi: 10.1101/gr.4695306.

[201] Alan Brelsford, Christophe Dufresnes, and Nicolas Perrin. Trans-species variation in Dmrt1 is associated with sex determination in four European tree-frog species. *Evolution*, 70(4):840–847, 2016. ISSN 15585646. doi: 10.1111/evo.12891.

[202] Martin Tesicky and Michal Vinkler. Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon? *Journal of Immunology Research*, 2015. ISSN 23147156. doi: 10.1155/2015/838035.

[203] Chen Su and Masatoshi Nei. Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. *Proceedings of the National Academy of Sciences of the United States of America*, 96(17):9710–9715, 1999. ISSN 00278424. doi: 10.1073/pnas.96.17.9710.

[204] P. J. Esteves, D. Lanning, N. Ferrand, K. L. Knight, S. K. Zhai, and W. Van Der Loo. The evolution of the immunoglobulin heavy chain variable region (IgV H) in Leporids: An unusual case of transspecies polymorphism. *Immunogenetics*, 57(11):874–882, 2005. ISSN 00937711. doi: 10.1007/s00251-005-0022-0.

[205] Fumi Miura, Kentaro Tsukamoto, Ratnesh Bhai Mehta, Kiyoshi Naruse, Wichian Magtoon, and Masaru Nonaka. Transspecies dimorphic allelic lineages of the proteasome subunit β-type 8 gene (PSMB8) in the teleost genus Oryzias. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50):21599–21604, 2010. ISSN 00278424. doi: 10.1073/pnas.1012881107.

[206] Kentaro Tsukamoto, Fumi Miura, Naoko T. Fujito, Goro Yoshizaki, and Masaru Nonaka. Long-lived dichotomous lineages of the proteasome subunit beta type 8 (PSMB8) gene surviving more than 500 million years as alleles or paralogs. *Molecular Biology and Evolution*, 29(10):3071–3079, 2012. ISSN 07374038. doi: 10.1093/molbev/mss113.

[207] Masaru Nonaka, Chisato Yamada-Namikawa, Martin F. Flajnik, and Louis Du Pasquier. Transspecies polymorphism of the major histocompatibility complex-encoded proteasome subunit LMP7 in an amphibian genus, Xenopus. *Immunogenetics*, 51(3):186–192, 2000. ISSN 00937711. doi: 10.1007/s002510050030.