No evidence for disassortative mating based on HLA genotype in a natural fertility population

Gillian L. Meeks[1,10], Brooke Scelza[2,10*], Katherine. M. Kichula[3], Catrinel Berevoescu[4], Kristin Hardy[5], Ticiana D.J. Farias[3], Genelle F. Harrison[3], Nicholas R. Pollock[3], Neus Font-Porterias[3,8], Sean Prall[2], Paul J. Norman[3,6,11**], Brenna M. Henn[5,7,8,9,11***]

[1]Graduate Program in Integrative Genetics and Genomics, University of California Davis, Davis, CA, USA, [2]Department of Anthropology, University of California Los Angeles, Los Angeles, CA, USA, [3]Department of Biomedical Informatics, University of Colorado School of Medicine, Aurora, CO, USA [4]Undergraduate Program in Genetics and Genomics, University of California Davis, Davis, CA, USA [5]Center for Population Biology, University of California Davis, Davis, CA, USA [6]Department of Immunology and Microbiology, University of Colorado School of Medicine, Aurora, CO, USA, [7]Department of Anthropology, University of California Davis, Davis, CA, USA [8]UC Davis Genome Center, University of California Davis, Davis, CA, USA

[9] Lead contact

[10,11]Indicates co-authorship position

[*]Correspondence: bscelza@gmail.com

[**]Correspondence: paul.norman@cuanschutz.edu

[***]Correspondence: bmhenn@ucdavis.edu

Summary

Evidence for disassortative mating based on the human-specific MHC, i.e. HLA, is equivocal[1]. Initial evidence for disassortative HLA-pairing in the European-descent Hutterites[2] has tended not to replicate in other populations. Recent studies, rather, reflect assortative mating associated with cosmopolitan population structure[1]. Although their configuration is more relevant to the majority of human evolutionary history, few small-scale, endogamous populations have been examined. Here, we test whether the extent of HLA dissimilarity between individuals influences mate choice by contrasting chosen partnerships (both “love” marriages and informal partnerships) with arranged partnerships among the Himba, an endogamous and formerly natural fertility group living in northern Namibia. With targeted sequencing we typed eight classical HLA genes for n=366 individuals and recorded self-reported partnership types and informal partnerships discovered through shared biological offspring. Across 249 known partnerships, we find no difference in HLA dissimilarity between chosen and arranged partnerships and no effect of the two partnership types that differs from a random, background distribution of partnerships. We also looked directly for fitness benefits of the potential offspring genotypes by testing HLA-based pathogen binding affinities, but also found no difference between arranged and chosen partners and no effects that differed from random partners. It is possible that due to high background HLA diversity in the Himba, most unrelated individuals in the population are sufficiently dissimilar at the HLA that there are few fitness benefits to offspring by maximizing mate dissimilarity. We do find, however, extensive haplotype sharing at the HLA region, suggestive of recent positive selection[3,4]. Episodes of fluctuating positive selection may be a more important driver in maintenance of HLA polymorphism than disassortative mating.

Results

The highly polymorphic major histocompatibility complex (MHC) region has a critical role in the vertebrate immune response, coding for molecules that bind foreign and self-peptides with specific affinities for recognition by T cells. MHC genes are codominantly expressed and therefore heterozygosity at each locus broadens peptide binding capabilities. Some vertebrate species, including nonhuman primates, perceive conspecific MHC molecules from body fluids through vomeronasal or olfactory neurons, and individuals tend to select mates with more dissimilar MHC genotypes. This suggests an evolved preference to increase MHC heterozygosity in offspring[1,5,6].

Experimental work in humans suggested that we too could discriminate HLA molecules through secretions, despite not possessing vomeronasal organs, with women tending to prefer the smell of men with more dissimilar HLA genotypes[7]. However, investigations comparing HLA dissimilarity among actual partnerships in the decades since have led to inconclusive results[1,8]. The only evidence for disassortative mating based on HLA genotype has been in two ethnically and genetically homogenous populations, the Hutterites[2] and Mormons[9]. In fact, in a recent meta-analysis the opposite trend, assortative mating at the HLA, was significant when considering ethnically heterogeneous study populations[1]. This latter finding likely reflects preferences for ethnic similarity in partners within a heterogeneous, cosmopolitan population, producing patterns of relatedness genome-wide, including at the HLA. The same meta-analysis showed that in experimental preference tests with ethnically heterogeneous study populations (e.g., odor tests, facial cues), as opposed to realized mate choice, preferences for HLA similarity were not found. These seemingly conflicting trends may reflect the importance of cultural norms, as well as phenotypic and behavioral preferences in mate choice, which could be overwhelming innate preferences for HLA dissimilarity. This would explain the conflicting patterns in observational mate choice versus experimental preferences[1].

There are several limitations to the observational mate choice data that is currently available. First, most HLA studies come from American and European populations. Of those that have been conducted outside these areas, we often know little about the accompanying sociocultural setting, including whether choice has been curtailed by processes like arranged marriage. In this study we aim to directly test the HLA dissimilarity hypothesis within Himba agropastoralists living in northern Namibia, combining genetics with ethnographic and demographic data that help to disentangle these factors. Several features of the Himba marital system make them a particularly useful population for studying mate choice effects on the HLA diversity. First, marriages can be either arranged (via parental choice) or chosen ("love matches"[10]). If there is an innate preference for HLA dissimilarity when selecting mates, we expect greater differences among love matches than among arranged partners. Additionally, concurrent marital and nonmarital partnerships are common among Himba, providing another avenue for active mate choice, even within the confines of an arranged marriage[11,12]. We use paired genetic (HLA and genome-wide) and partnership status to investigate whether there is greater HLA dissimilarity in unions that result from partner choice than in those that are arranged. We also examine the pathogen binding ability of hypothetical offspring of arranged and chosen partnerships to test for any direct fitness benefits of parental HLA allele combinations.

HLA diversity in the Himba has not been previously characterized. With HLA allele calls from 366 individuals, we assessed the diversity of HLA alleles based on their P group designation, that is, alleles were grouped according to identical peptide binding domains to indicate differences in pathogen binding function between alleles, across three HLA class I (A, B, C) and five HLA class II genes (DPα1, DPβ1, DQα1, DQβ1, DRβ1) (Supplementary Figure 1). We compared Himba HLA diversity to other publicly available populations curated to represent a diverse world-wide sample[1] and all available sub-Saharan African groups[13]. In order to make unbiased comparisons of allelic diversity based on sample size, we subset our sample to the maximum set of unrelated Himba individuals (n=102, maximum 4th degree relations) (Supplementary Figure 2). Despite endogamy and a recent population bottleneck[14], the allelic

diversity at each gene falls along the linear prediction based on sample size constructed from the publicly available data, with expected or somewhat greater diversity than expected given the sample size compared to other African populations (Supplementary Figure 3). We ensured genotype frequencies at each HLA gene did not deviate from Hardy Weinberg equilibrium (Supplementary Table 1) as this could indicate genotyping error or population substructure[15].

In order to address whether HLA dissimilarity differs between chosen and arranged partnerships, we first assessed whether there was an effect of partnership type on overall genetic relatedness. If mate choice is, in part, driven by minimizing overall genetic relatedness to reduce fitness consequences associated with increased homozygosity in offspring[14,16,17], chosen partnerships could be more genetically dissimilar overall than arranged partnerships. This could produce an associated, but spurious pattern of HLA dissimilarity in chosen partnerships as compared to arranged partners and the background distribution. On the other hand, preferences for cross-cousin marriage and strong ties between allied families could lead to greater sharing of IBD segments among arranged marriage partners than a background distribution would predict leading to associated, but spurious patterns of greater HLA similarity in arranged partners as compared to chosen partners and the background distribution.

We assessed overall relatedness via the total number of centimorgans shared in genome-wide IBD segments between the individuals in the partnership (see eq 2 in *Methods*). There was no effect of arranged partnership or chosen partnership on genome-wide IBD sharing, relative to a background population distribution created from random, unrelated (maximum 3rd degree relations), heterosexual pairings with realistic age gaps (Figure 1A) (p = 0.84, 0.82, for arranged and chosen partnership effects, respectively) nor a significant difference between the effect of arranged versus chosen partnership in a pairwise contrast of effect size estimates (p = 0.93). 3rd degree relatives were allowed in the background distribution to reflect the stated cultural preference for cousin marriages in arranged unions[10,18].

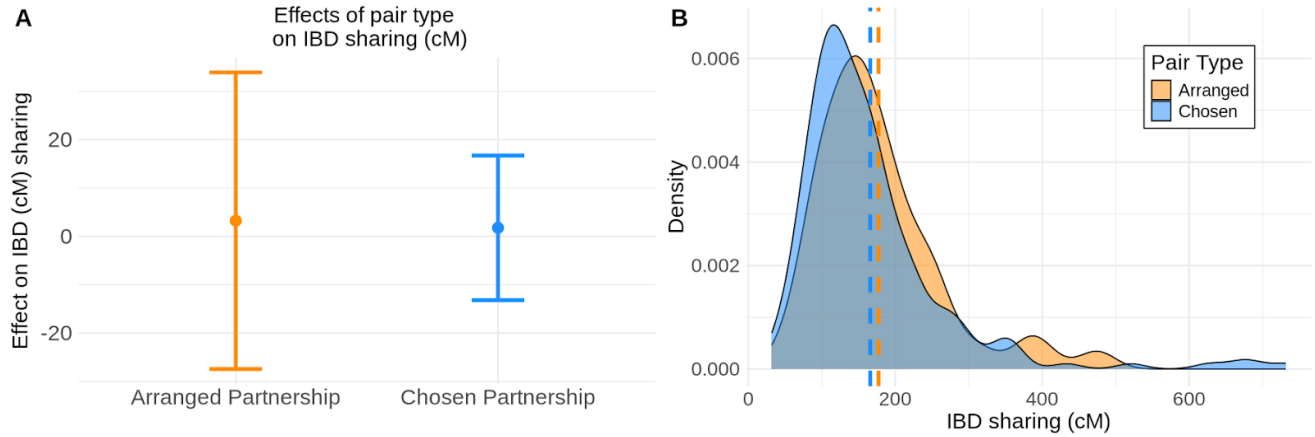

Figure 1: *No effect of partnership types on genome-wide IBD sharing*. A) We consider a linear model where the intercept is the expected genome-wide IBD sharing based on a background distribution of unrelated female - male pairs from the population. We plot the effect size estimates of arranged partnerships and chosen partnerships on IBD sharing and find the 95% confidence intervals of the effects do not significantly differ from 0. There is no significant effect of arranged or chosen partnerships on IBD sharing. B) The distributions of the IBD sharing (cM) for arranged and chosen partnerships clearly overlap. Means of each distribution are noted with a dashed line.

HLA similarity does not differ among partnership types

We next addressed our main question: do chosen partnerships have lower HLA similarity, after accounting for their overall relatedness, than arranged partnerships? That is, is mate choice driven by minimizing relatedness at the HLA region rather than genome-wide relatedness? We assessed HLA

similarity in two ways, first by constructing a HLA similarity score, the number of shared alleles between a pair of individuals across all 8 genes (i.e. minimum score of 0, maximum score of 16) and also by computing the average log odds of homozygous offspring across all 8 genes (see *Methods*). These two metrics are highly correlated (Pearson r = 0.96), but average log odds of homozygous offspring provides a finer measurement (Supplementary Figure 4). Using linear mixed models that correct for overall IBD sharing and account for the influence of each female and male's HLA genotypes to their similarity scores across all pairings (see eq 3 in *Methods*), we found no effect of arranged or chosen partnership on overall HLA similarity (p = 0.7, 0.61, respectively) (Figure 2A) or on average log odds of HLA homozygosity in potential offspring (p = 0.79, 0.80, respectively) (Figure 2C). In a pairwise contrast there were no differences in the effect of chosen versus arranged partnership effect on HLA similarity score (p = .565) or on average log odds of HLA homozygosity in potential offspring (p = .72).

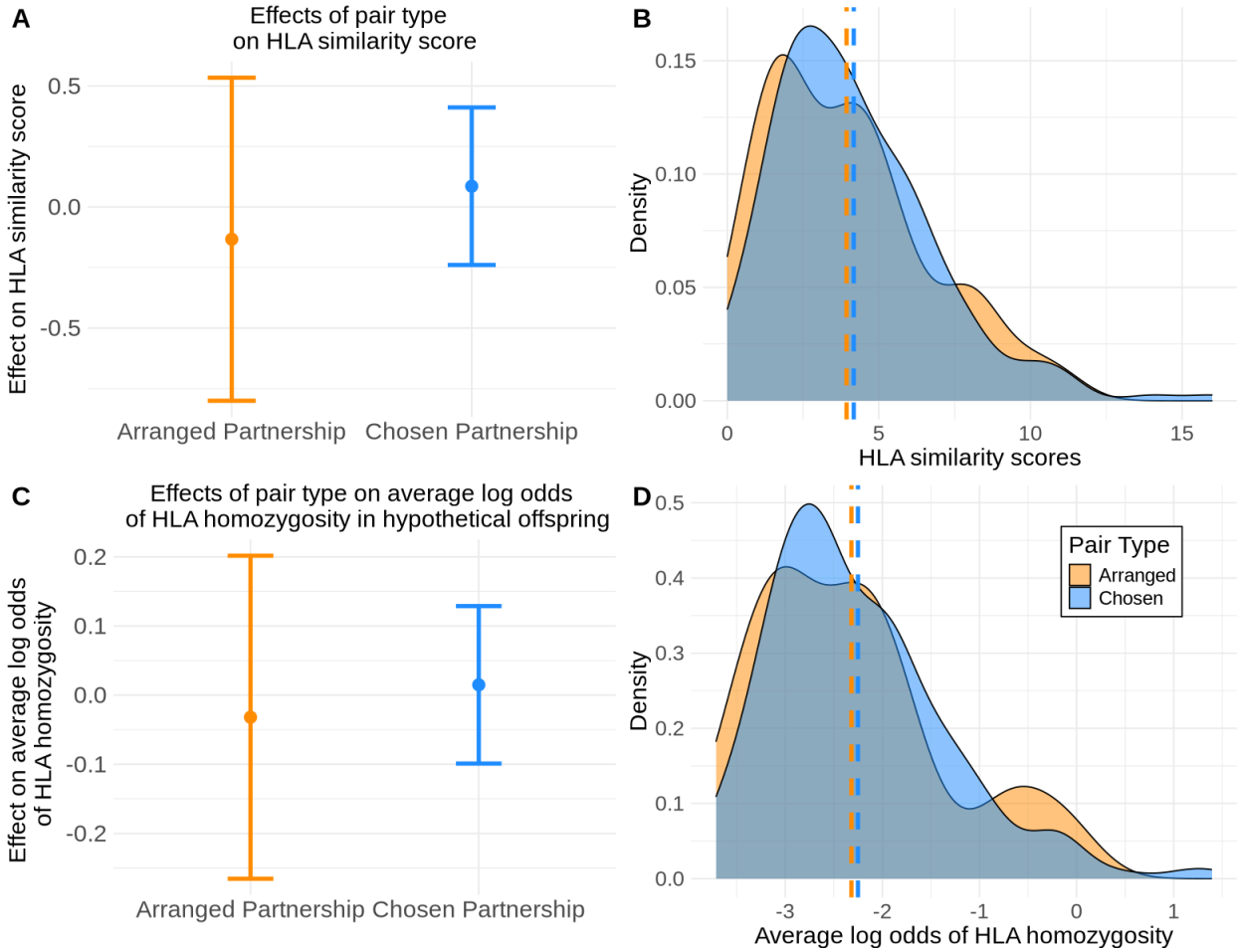

Figure 2: *No effect of partnership types on HLA dissimilarity or average log odds of offspring HLA homozygosity.* We again consider linear models where the intercept is the expected outcome based on a background distribution of random, unrelated, opposite sex pairs from the population and control for genome-wide IBD sharing. We apply the model to the HLA dissimilarity score (A) and to the average log odds of HLA homozygous offspring (C) across all 8 genes. We plot the effect size estimates of arranged partnerships and chosen partnerships on HLA similarity score (A) and average log odds of offspring homozygosity (C) and find the 95% confidence intervals of the effects do not significantly differ from 0. The distributions of HLA similarity scores (C) and average log odds of offspring HLA homozygosity (D)

clearly overlap between arranged and chosen partnerships. Means of each distribution are noted with a dashed line.

Pathogen binding repertoires of potential offspring do not differ among partnership types

Increased pathogen binding ability of offspring is theorized to be the alleged evolutionary driver of disassortative mating at the MHC region[19,6]. Therefore, even if similarity scores, as assessed by number of shared alleles and average odds of offspring homozygosity across the HLA do not differ among pair types, individuals might seek out complementary HLA genotypes in mates to increase the pathogen binding ability in offspring. Using the same linear mixed model (*Methods* eq 3) as above, we looked for effects of arranged and chosen partnerships on the average computationally-predicted pathogen binding repertoires in hypothetical offspring (i.e., the average number of unique pathogen peptides predicted to be bound by each offspring HLA genotype). We used a curated list of 25 pathogens, taking a subset of those listed in a previous study of evolutionarily significant human pathogens[20] that have plausible transmission in the Himba, and included additional sexually transmitted pathogens known to be prevalent in the Himba[21–23] (personal correspondence with Dr. Mary Ashley Hazel) (Appendix 1). The HLA alleles from our sample were run through netMHCpan[24] and netMHCIIpan[25] to predict the binding affinities of each allele to our list of antigenic pathogen peptides. At each locus, the hypothetical offspring genotypes, in expected proportions, were cross-referenced to allele binding affinities to produce a mean number of unique pathogen peptides that bind with high affinity per HLA genotype. The HLA class II gene product pairs, DPα1 and DPβ1, and DQα1 and DQβ1, can each form cis and trans heterodimers, comprising DP and DQ epitopes, respectively[26,27]. Thus, an individual heterozygous for both DPα1 and DPβ1 can form 4 unique DPα1/DPβ1 molecules. We assessed the mean number of bound antigenic peptides for all possible offspring DP and DQ epitope combinations, again based on expected genotype proportions.

Across all 8 HLA genes tested, there was no effect of arranged or chosen partnerships on the mean number of pathogens bound by offspring genotypes that differed from the background distribution of random partners (Supplementary Figure 5, Supplementary Table 2). Likewise, we found no differences in the effect estimates of arranged and chosen partnerships in pairwise contrasts (Supplementary Figure 5, p-values in Supplementary Table 3). HLA class I molecules generally bind peptides derived from intracellular pathogens (e.g. viruses) whereas HLA class II molecules generally bind those derived from extracellular pathogens (e.g. bacteria and parasites)[28] . Thus, we also performed this analysis including only intracellular pathogens for HLA-A, B, and C genotypes and only extracellular pathogens for HLA-DP, DQ, and DRβ1 genotypes. We still found no significant effects of arranged or chosen partnerships in mean pathogens bound and no significant differences of effects in pairwise contrasts (Supplementary Figure 6, Supplementary Tables 4-5).

Extensive haplotype sharing could indicate recent episodes of fluctuating positive selection

We find no evidence that disassortative mating is driving maintenance of polymorphism at the HLA. Other forms of selection, besides sexually mediated balancing selection, could therefore play a larger role in maintaining polymorphism at this locus. We assessed haplotype sharing across the entire 5Mb HLA genomic region by visualizing the SNP-based haplotypes (from genotyping array) across a 5Mb region surrounding the 8 typed HLA genes. The SNP-based haplotypes along this much larger 5Mb region were grouped by their HLA allele haplotypes (typed via short-read sequencing) (Figure 4) (see *Methods* haplotype resolution). Although the 7 HLA genes used to create allele-based haplotypes comprise only ~1% of the base pairs in this 5Mb region, we see virtually identical SNP haplotypes across unrelated individuals (maximum relatedness of 4th degree) who share frequent HLA allele-based haplotypes (n=92 haplotypes with >4 occurrences in the 102 unrelated individuals) (Figure 4, Supplementary Figure 7). This finding indicates that the entire 5Mb HLA region was inherited without recombination in these unrelated individuals, consistent with recent positive selection on these haplotypes[3,4].

Unlike our previous study of sub-Saharan African populations, where in 6 of the 7 populations the majority of haplotypes were observed only once[13], 70% of haplotypes were observed at least twice in our set of unrelated individuals, with the most frequent haplotype (A*30:02 - C*03:04 - B*15:10 - DRβ1*03:01 - DQα1*05:01 - DQβ1*02:01 - DPβ1*01:01) observed 23 times in the unrelated set. Nineteen of the 29 haplotypes occurring more than once in the unrelated set of individuals were observed in at least one of the 7 sub-Saharan African populations from our previous study (Supplementary Figure 8). We assessed whether there was a relationship between haplotype frequency and number of unique peptides computationally predicted to be bound by each haplotype[29], but found no significant relationship (p = 0.91, Supplementary Figure 9) and likewise found no significant relationship with number of peptides bound when we analyzed each pathogen independently (Supplementary Figure 10).

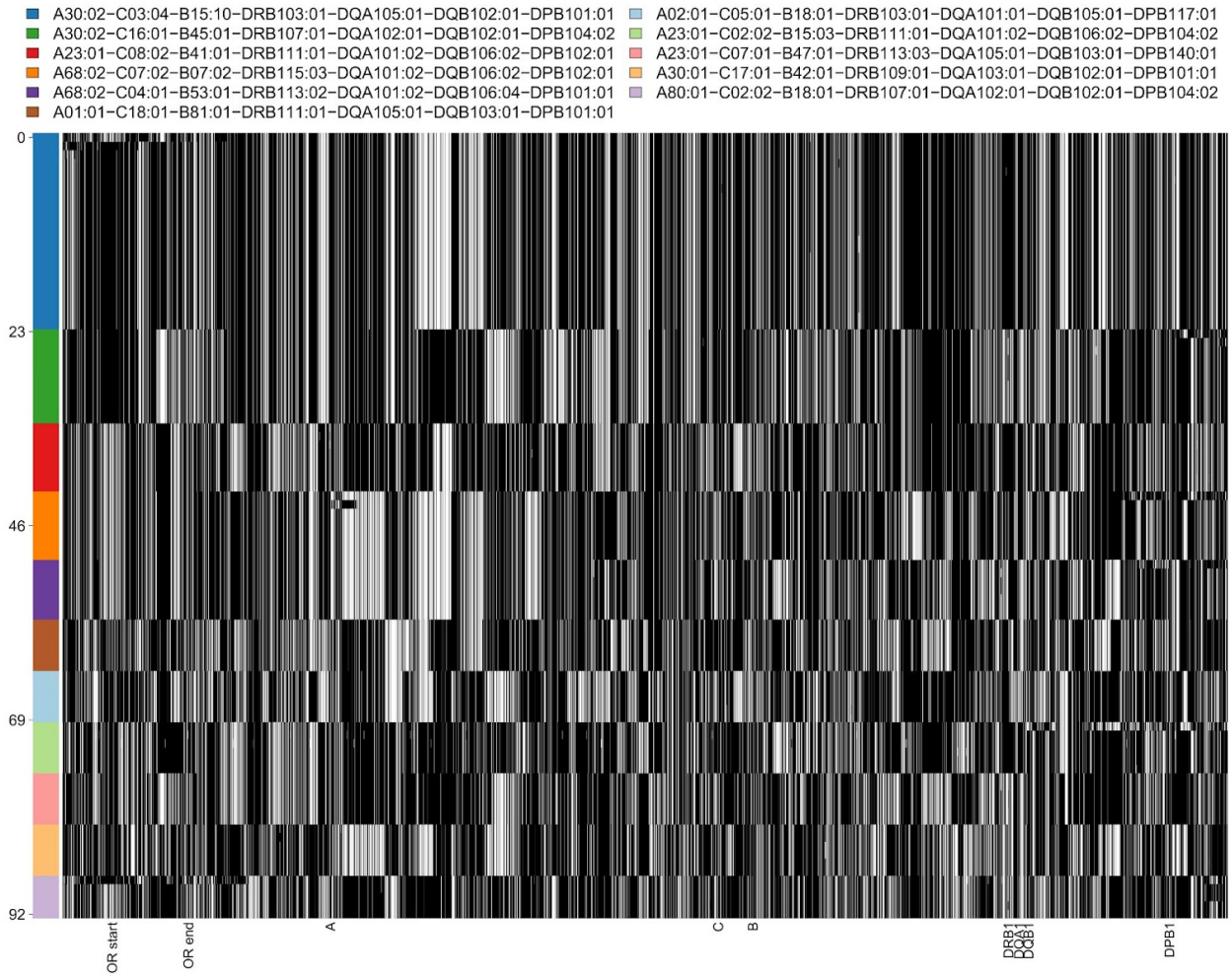

Figure 3: *Extensive haplotype sharing across the 5 Mb HLA region among unrelated individuals*. SNP-based haplotypes (rows) are grouped according to HLA-allele based haplotypes (for those haplotypes with > 4 occurrences in the 102 unrelated individuals). Columns indicate genomic positions with white/black designating the ref vs. alt allele at each position. The haplotypes are grouped by their allele calls at HLA-A - C - B - DRβ1 - DQα1 - DQβ1 - DPβ1. We were unable to resolve the haplotypes for DPα1 alleles calls. Haplotypes were visualized with haplostrips[30]. The approximate, relative locations of the 7 HLA genes are shown on the x-axis along with the approximate starting and ending positions of the upstream Olfactory Receptor (OR) gene cluster.

Consistent with the haplotype visualization, we found increased IBD sharing at the HLA region relative to genome-wide levels. This pattern implies recent (within 200 generations) positive selection as opposed to overdominance (i.e., heterozygote advantage)[3,4]. The mean rate of IBD sharing across the HLA region between random pairs of unrelated individuals (the background distribution for our pair analyses; n=352 individuals, n=18,487 pairs) is 5.3%, 1.97 standard deviations above the genome-wide median of 3.1%. The rate of IBD sharing across the HLA region in the subset of unrelated individuals who have common haplotypes (n=71 individuals with common haplotypes, those shown in figure 3) is even higher at 10.4%, 6.4 standard deviations above the genome-wide median of 2.8%.

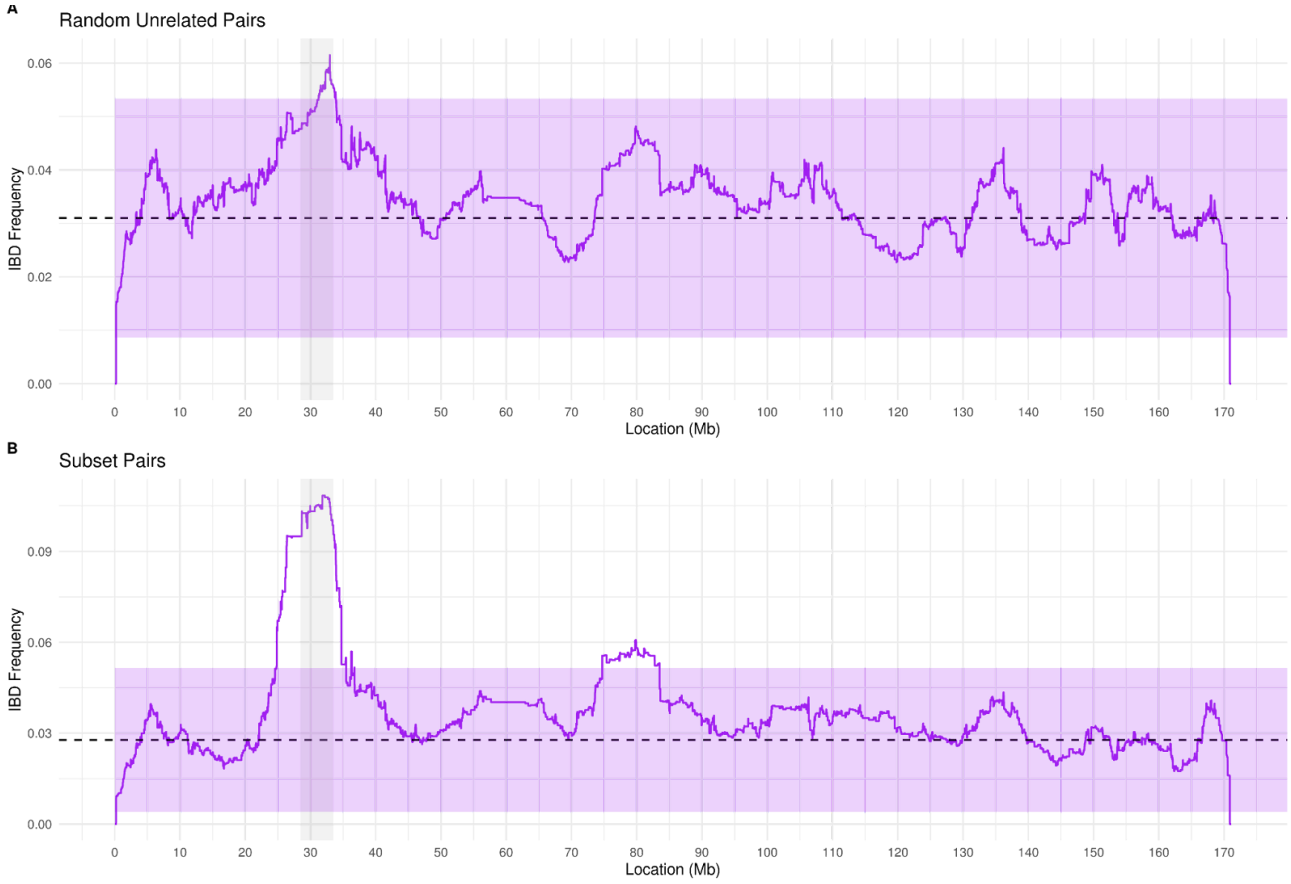

Figure 4: *Elevated IBD sharing rates at the HLA region is consistent with recent, fluctuating positive selection.* Panels A and B depict IBD sharing rates across chromosome 6 (rates were averaged across 1000bp bins). The dashed line depicts the genome-wide median IBD sharing rate, excluding chromosome 6. Panel A depicts IBD sharing rates for the pairs that make up the background distribution for the pair analyses (random heterosexual pairings, with relevant age gaps, related at no closer than 3rd degree). Panel B depicts the IBD sharing rates for all the unrelated individuals included in the allele frequency estimations (no closer than 4th degree relatives) who have the common haplotypes shown in Figure 3 (n=71 individuals). The HLA region is designated in grey. The purple shaded region depicts +2 and -2 standard deviations from the genome-wide (excluding chromosome 6) median IBD sharing rate, 3.1% and 2.8% for panel A and B, respectively.

Discussion

There is evidence for MHC-based mating preferences in a range of species, including non-human primates, but evidence in humans is equivocal[1]. Whereas some experimental work has found evidence of innate ability to discriminate HLA dissimilar mates through body odours, realized mate choice studies are inconsistent, with some showing preferences for similarity, others for dissimilarity, and others, like ours, showing no preference[1]. Our study is free from the limitations of cosmopolitan samples that exhibit spurious patterns of HLA similarity in realized partners due to preferences for ethnically, and therefore, relatively genetically similar partners. Cultural norms like arranged marriage and cross-cousin marriage preferences are predicted to promote similarity across the HLA, while tolerance of extramarital partnerships and "love match" marriages could enhance the possibility of finding a HLA-dissimilar partner. Therefore, the unique socioecology of Himba marriage patterns sets up a strong test of the dissimilarity hypothesis.

Surprisingly, we find that chosen partnerships are neither less similar at their HLA than random pairs constructed from a relevant background distribution, nor do they differ in their HLA similarity from arranged partnerships. Likewise, we find no evidence that chosen partners are selected to directly maximize offspring pathogen binding repertoire. Unlike the Hutterites[2] and Mormons[9] that do show evidence for greater HLA dissimilarity in married pairs, background genetic diversity in the Himba is likely high enough that any unrelated individual is sufficiently dissimilar at the HLA and there may be no fitness benefit to maximizing dissimilarity. Furthermore, it has been suggested that offspring may not benefit from maximal diversity at the HLA; T cells that bind self peptide-MHC complexes too strongly undergo pruning in the thymus and possessing many different MHC alleles could deplete the T cell repertoire required for appropriate immune response.[31,32] This process of negative selection for extremely autoreactive T-cells has been hypothesized to inhibit the expansion of the MHC genomic region to include additional copy numbers.[33] Supporting this optimality hypothesis, decreased parasite load in intermediate versus maximally diverse MHC individuals has been observed in voles[32] and stickleback fish.[34]

Given the limited benefits of HLA dissimilarity in this context, assortative mating for sociocultural traits may be a more important driver of mate choice for groups like the Himba than genetic compatibility. Our previous work has shown that resource scarcity affects women's partner preferences[35,36], and that relationship success is correlated with similarity in mate value[37]. Concurrent partnerships, rather than increasing the chance of finding a partner with a dissimilar HLA genotype, may be more important for garnering resource security for women, and distributing reproductive opportunities for men who are otherwise limited by polygyny and a later age at marriage[11,38].

We find extensive IBD sharing at the HLA region consistent with recent, fluctuating, positive selection rather than heterozygote advantage maintaining polymorphism at the region[3,4]. Despite association with many diseases[39] and consistently showing up as a target of natural selection[29,33], the selective forces shaping HLA diversity are not well understood[33]. We suggest that innate drive for disassortative mating based on HLA genotype is unlikely to be a major evolutionary force in maintaining diversity and future work should focus on assessing the contribution of other selective regimes (e.g., negative frequency dependent, fluctuating selection, heterozygote advantage, divergent allele advantage, etc.) to the maintenance of polymorphism at the HLA region.

Resource Availability

Lead contact: Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Brenna M. Henn (bmhenn@ucdavis.edu)

- **Materials Availability:** This study did not generate new unique reagents.
- **Data and code availability:** Genotype data for the individuals included in this study have been previously deposited at dbGap as phs001995.v2. HLA allele calls will be deposited at this dbGap accession upon publication. All original code used to conduct analyses will be deposited in the following github repository https://github.com/gillianmeeks/Himba_HLA upon publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgements

We thank the Himba communities in which we have sampled; without their support, this study would not have been possible. We thank Mark Grote and Cole Williams for assistance with statistics and data analysis. BMH and GLM were supported by the National Institutes of Health (NIH) grant R35GM133531. BS was supported by the National Science Foundation grant BCS-1534682. BMH, GLM, TDJF, and P.J.N. were supported by NIH award R01AI151549. The authors acknowledge the High Performance Computing Core Facility at the University of California, Davis for providing computational resources that have contributed to the research results reported in this paper. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation. The funders had no role in the decision to publish or prepare the manuscript.

Author Contributions

BS, PN, and BMH conceived of the study. GLM conducted analyses with help from KH and CB. GLM and BS wrote the manuscript with input from PN, SP, and BMH. KK, GH, NFP, TDJF, and NP generated HLA sequence data under the supervision of PN. Sample and ethnographic data collection was conducted by BS and SP. BMH supervised the research.

Declaration of Interests

The authors declare no competing interests.

Supplemental information

Document S1. Figures S1–S12 and Tables 1-6

Appendix 1. Spreadsheet listing pathogen peptide NCBI accessions

Methods

Population Description

The Himba reside in the Kunene Region of Northwestern Namibia and Southern Angola, but our study was conducted specifically in the community of Omuhonga. The Himba are a seminomadic, agro-pastoralist, non market-integrated population, who are closely related to the Herero.

Data Generation

DNA was collected with informed consent from individuals via saliva using the prepIT-L2P kit and protocol. DNA was used to generate HLA genotype calls for 366 individuals. HLA class I and II loci were targeted for DNA sequencing using a well-established biotinylated DNA probe-based capture method[40].

HLA alleles were determined from the sequence capture using the consensus calls obtained from two algorithms: NGSengine® 2.10.0 (GenDX, Utrecht, the Netherlands) and HLA Explore™ (Omixon Biocomputing Ltd. Budapest, Hungary). Alleles were then converted into their P-group designations from https://hla.alleles.org as of 10/7/2024 (2-field designation) which groups alleles whose nucleotide sequences encode identical protein sequence in the peptide binding domains (exon 2 and 3 for HLA class I and exon 2 only for HLA class II alleles). Genotype array data using one of two arrays, Illumina H3Africa or MEGAex, were previously generated for 360 of these individuals [dbGaP: phs001995.v2.]. One individual was newly genotyped using H3Africa; five individuals with HLA capture data did not have corresponding SNP array data and were not included in the partnership analyses requiring genome-wide relatedness metrics. These SNP array datasets previously underwent quality control measures and were merged as described[12]. There were 219 individuals (82 males, 137 females) having both genotype array and HLA genotype data who were in at least one documented partnership: arranged marriage, love match marriage, self-reported informal partnership, or an informal partnership. The latter were discovered due to shared biological offspring as determined by a list of parent offspring trios generated via KING 2.1.3[41] with the identical by descent (IBD) segment flag. There were 47 arranged marriages, 47 love match marriage pairs, and 155 informal pairs. Of the informal pairs, 48 were discovered only through shared biological offspring. These unreported informal partnerships could be due to self-report bias. The 47 love match marriages were combined with 203 informal partnerships and designated as chosen partnerships to compare with the 47 arranged partnerships in our analyses.

Genotype array data

Individuals were genotyped using one of two arrays, H3Africa or MEGAex. These datasets previously underwent quality control measures and were filtered using PLINK v1.90b3v[42] for missingness greater than 5%, a MAF less than or equal to 1%, and a Hardy–Weinberg equilibrium exact test with a *p*-value below 0.0001 as described previously[12] and documented in dbGaP phs001995.v2. The merged array dataset containing 376,657 SNPs was used for analyses.

Hardy-Weinberg Analysis

To test whether the HLA genotypes complied with Hardy-Weinberg equilibrium, we used the Asymptotic Statistical Test with Ambiguity (ASTA) from the HWETests package[43]. ASTA is designed to evaluate deviations from Hardy-Weinberg equilibrium for highly polymorphic loci while accounting for ambiguity in genotype calls. Using ASTA, we ran the full_algorithm() function with default parameters on a set of 102 unrelated individuals (maximum relation of 4th degree).

Identity by Descent (IBD) Sharing

Genotype data was phased using SHAPEIT v2.r837[44] with --duohmm -W 5 using an ancestry-matched recombination map[45], constructed based on linkage disequilibrium patterns in the 1000 Genomes Project[46] Yorubans, matching the predominant West African ancestry present in the Himba. We called IBD segments with phasedIBD[47] using default parameters with phase correction turned off. We determined phasedIBD calls to be slightly more accurate than germline 1.5.3[48] and GERMLINE2[49] by comparing the total cM of IBD shared between parent offspring pairs (Supplementary Table 6). We expected parents and offspring to share the full length of the recombination map in IBD1 and share an additional background amount of IBD2 estimated to be the average cM of IBD shared between unrelated (> 3rd degree relations) individuals in the population. The phasedIBD calls without phase correction most closely matched this expectation. The phasedIBD calls with phase correction option turned on were the least accurate. A minimum threshold of 3cM was used for all runs.

Creating a set unrelated sets of individuals

Traditional pedigree inference methods that use expected IBD sharing perform poorly in endogamous populations with elevated background sharing and pedigree reticulations. We used PONDEROSA[50] recently developed in our lab, to accurately infer close relationships in our sample. PONDEROSA uses a machine learning approach to learn the population distribution of IBD summary statistics to more accurately classify close relationships. Based on the inferred relations we excluded any pairs who were 2nd degree or closer relatives from our background distributions in our linear models (eq 2-3). Third degree relative pairs were permitted in this background distribution to reflect the stated cultural preference for cousin marriages in arranged unions[10,18]. We also used this program to maximize a set of unrelated individuals for allele and haplotype frequency based analyses (n=102 individuals with maximum 4th degree relations).

HLA similarity score

We assessed HLA similarity in terms of the number of alleles shared across the eight typed HLA genes for each pair (minimum of 0 if no alleles are shared at any gene, maximum of 16 if both alleles at each gene are shared). A shared allele in a homozygous state in one individual only counted as one shared allele between the individuals. Importantly, due to the many alleles at each HLA gene, two individuals with heterozygous genotypes at a locus does not necessarily imply shared alleles.

Average log odds of homozygous offspring

Odds of homozygous offspring for each locus were determined according to standard mendelian transmission probabilities, accounting for the multi-allelic states of each gene. That is, due to the many alleles at each gene, if both individuals in a pair are heterozygous that does not mean they have the same genotype. Table 2 shows the possible states for this variable considering only one gene. Importantly, for a given number of alleles shared at a locus, there are multiple possibilities for the odds of homozygous offspring depending on the genotype states of the individuals in the pair. We log transformed these odds to normalize the distribution and then averaged across the 8 loci. We included a +0.1 boundary correction in the numerator and denominator of the odds calculations to allow for numerators and denominators of 0 (eq 1).

Table 2: Possible states of odds of homozygous offspring at one locus

Number of shared alleles by pair	Partner 1 genotype	Partner 2 genotype	Odds of homozygous offspring $\left(\frac{\#\ events\ homozygous}{\#\ events\ heterozygous}\right)$
2	Homozygous (e.g., A1/A1)	Homozygous (e.g., A1/A1)	$\frac{4}{0}$
2	Heterozygous (e.g., A1/A3)	Heterozygous (e.g., A1/A3)	$\frac{2}{2}$
1	Homozygous (e.g., A1/A1)	Heterozygous (e.g., A1/A3)	$\frac{2}{2}$
1	Heterozygous (e.g., A1/A3)	Heterozygous (e.g., A1/A4)	$\frac{1}{4}$
0	Homozygous or Heterozygous (e.g., A1/A1 or A1/A4)	Homozygous or Heterozygous (e.g., A3/A3 or A3/A5)	$\frac{0}{4}$

For a given number of shared alleles (1 or 2) there are multiple possible odds of homozygous offspring depending on if the individuals in the pair are heterozygous or homozygous at the locus.

$$eq\ (1)\ \ Average\ log\ odds\ of\ homozygosity\ = \frac{\sum_{i=1}^{8} log(Odds_i)}{8} \quad \text{where} \quad Odds_i = \{ \frac{0+0.1}{4+0.1}, \frac{1+0.1}{3+0.1}, \frac{2+0.1}{2+0.1}, \frac{4+0.1}{0+0.1} \}$$

Pathogen Proteins

We modified a previously published list of representative human pathogen proteins[20] that was curated from the Gideon database[51] based on global distributions, potential for high mortality, and historical relevance. After personal correspondence with an infectious disease expert, Dr. Ashley Hazel, with extensive experience working with the Himba community we removed 6 pathogens unlikely to be ecologically relevant in Northern Namibia: *Entamoeba histolytica* (amebiasis), *Trichinella spiralis* (roundworm), *Schitosoma mansoni* (schistosomiasis), HIV, Hepatitis C, and *Yersinia pestis* (plague). We added 4 sexually transmitted diseases known to be highly prevalent in the Himba (HSV1, HSV2, *Chlamydia trachomatis* (chlamydia), and *Neisseria gonorrhoeae* (gonorrhea)) for a total of 25 pathogens[52]. We used the NCBI accession numbers available in the previously published study to gather amino acid sequences for the antigenic proteins of the 21 pathogens overlapping that publication. For the 4 additional pathogens we used IEDB https://www.iedb.org/ to find the pathogen proteins that are MHC ligands and then identified these proteins' amino acid sequences using NCBI. The complete list of 298 proteins and their accession numbers are available in Appendix 1.

Pathogen Binding Prediction

We used NetMHCpan (v4.1)[24] to predict peptide binding affinity for the alleles of HLA class I genes (A, B, and C) and used NetMHCIIpan (v4.3)[25] for the alleles of HLA class II genes (DPα1, DPβ1, DQα1, DQβ1, DRβ1). These algorithms are trained using experimental binding affinity and eluted ligand data to predict binding affinity. Binding prediction was performed on all possible amino acid 9mers to match the 9 residue binding groove for class I genes and all possible amino acid 15mers for class II genes. The class II genes with two subunits typed, 1) DPα1 and DPβ1, 2) DQα1 and DQβ1, combine molecules to form the DP and DB epitopes and thus all possible combined DP and DQ molecules were cross-referenced for their binding affinities. The predicted binding affinity scores between every possible peptide and every MHC molecule are ranked as a percentage by the algorithms based on comparison with a large set of naturally occurring peptides. We used the default rank threshold for strong binding peptides (top 0.5% for class I genes, and top 1% for class II genes). One of the rare alleles (DPβ1*61:01N) was not found in the netMHCIIpan database because it is a null allele, i.e., does not present peptides, and thus pairs including the individual with this allele were excluded for DP binding analysis.

Statistical Models

We constructed a model to assess the effect of partnership type on IBD sharing (eq 2) and on our HLA metrics (eq 3): HLA similarity score, mean log odds of homozygous offspring, and mean number of predicted pathogens bound by hypothetical offspring genotypes. Each partnership type (arranged, chosen) was encoded as a binary variable, and random effects were included for each female and each male to account for individual-level variation. IBD sharing was included as a covariate in equation 3 to determine whether differences in HLA metrics between partnership types remained significant after conditioning on overall genetic relatedness.

The intercepts of both models represent a baseline derived from all 18,487 possible opposite sex pairings that are not in known partnerships, are no closer than 3rd degree relatives, and also constrained to those that have a permissible age differential based on the range observed in the known partnership

data. This baseline provides a background distribution for the expected similarity of partners in the population if there is random mating, but with avoidance of close kin. We allowed 3rd degree relatives to exist in our background distribution of partners as Himba individuals report a cultural preference for cousin marriages in arranged unions[10]. However, surprisingly, no individuals in actual arranged partnerships were related at 3rd degree or closer but 1 “love” marriage and 2 informal partnerships discovered through pedigree analysis were found to be 3rd degree relatives.

This background distribution enables testing whether each pair type deviates from the baseline expectation of similarity while providing additional information to estimate the random effects. The random effects of each female and male capture the unique contribution of an individual's genotype (at the HLA or genome-wide for the IBD analysis) to their similarity scores across all pairings (both real and random). This approach allows us to identify whether individuals’ real partnerships (arranged and chosen) differ systematically in their similarity scores when compared to their average similarity with potential random partners in the population.

$$(eq\ 2) \qquad IBD = 1 + \beta_{arranged\ partnership} + \beta_{chosen\ partnership} + \beta_{(1|female\ ID)} + \beta_{(1|male\ ID)}$$

$$(eq\ 3) \qquad HLA\ metric = 1 + \beta_{IBD} + \beta_{arranged\ partnership} + \beta_{chosen\ partnership} + \beta_{(1|female\ ID)} + \beta_{(1|male\ ID)}$$

Pairwise contrasts in the effect size estimates of the two different partnership types (arranged vs chosen) were assessed by calculating the Z-statistic, defined as the contrast estimate divided by its standard error. P-values for these Z-statistics were derived from a two-tailed normal distribution. Statistical significance for the main effects and pairwise contrasts was determined using a p-value threshold of 0.05.

HLA haplotype resolution

The HLA allele calls from target sequence capture are unphased, but we were able to resolve HLA haplotypes by imputing closely matched HLA allele calls from the paired, phased SNP data across the 5Mb region encompassing the typed HLA genes. Although HLA imputation based on sparse genotype data is not accurate at detailed resolutions, we were still able to resolve the haplotypes by comparing imputed alleles to the known alleles from the HLA targeted sequencing. We trained HIBAG HLA imputation classifier models[53] on the Himba SNP data across the whole 5Mb region with paired HLA allele calls from the target capture (n=361 individuals). The models had “out of bag” accuracies of 97.96%, 96.97%, 98.60%, 98.75%, 98.73%, 98.75%, 97.66% across A, B, C, DPβ1, DQα1, DQβ1, and DRβ1, respectively. The program does not impute DPα1 genotypes. We then extracted each phased SNP-based haplotype at the HLA region and created two pseudo-homozygous genotypes for each individual (i.e., ID1_hap1|ID1_hap1 and ID1_hap2|ID1_hap2). We ran these pseudo-homozygous genotypes through our curated HIBAG model to predict the HLA allele associated with each pseudo-homozygous genotype. We ensured that the predicted alleles across the two pseudo-homozygous genotypes matched one of the known target sequenced HLA allele calls for at least the first field. In this way, we filled in the allele calls across the two phased haplotypes. We were able to resolve 710/722 haplotypes with this method. The linkage disequilibrium patterns determined via asymmetric conditional linkage disequilibrium closely correspond to previously identified patterns[54] with high LD between HLA-B and C and also among HLA-DRβ1, DQα1, and DQβ1, demonstrating the accuracy of our haplotype resolution (Supplementary Figure 11).

Calculating IBD sharing rates

To calculate IBD sharing rates at each base pair genome-wide, a Line Sweep algorithm was developed in Python (See Supplemental Figure 12). In this method, the start of an IBD segment is marked +1, and the base pair directly following the last base pair in the IBD segment, is marked -1. C) A ‘line’ sweeps across

the chromosome-length array summing the tally at the current base pair with the tally carried over from the previous base pair to generate IBD coverage scores for each base pair. In this way, a base pair located within a larger set of IBD fragment markings receives a larger coverage value during the line sweep step. The IBD sharing rate for each base pair is then computed by dividing the IBD coverage score by the total number of possible haplotypes in the dataset. The number of possible haplotypes within a dataset is determined as the number of pairs (individuals choose 2) multiplied by 2. This procedure reduces the complexity of the IBD sharing calculation problem, allowing for efficient computation of the IBD sharing rate at each base pair, genome-wide. IBD sharing rates were binned and averaged in 1000 base pair chunks. These IBD coverage and proportion calculation programs are available on the GitHub repository.

Ethics

Ethical approval for this study was granted by the University of California, Los Angeles (IRB-10-000238), the State University of New York, Stony Brook (IRB-636415-12), and was approved by the Namibian Ministry of Home Affairs and the University of Namibia Office of Academic Affairs and Research. Chief Basekama Ngombe provided permission to work in the community and local approval of the study. Community leaders were actively involved in discussions regarding what genetic data could be used for, who would have access to it, and whether there was a for-profit element involved (there was not). Individual informed consent, and for minors' parental assent, was obtained orally from all participants, given low rates of literacy in the community. Care was taken to protect participants' privacy, for example via a double-blind procedure for DNA collection[55]. These data were collected as part of the longitudinal Kunene Rural Health and Demography Project, which has been working in the community since 2010.

References

1. Winternitz, J., Abbate, J. L., Huchard, E., Havlíček, J. & Garamszegi, L. Z. Patterns of MHC-dependent mate selection in humans and nonhuman primates: a meta-analysis. *Mol. Ecol.* **26,** 668–688 (2017).
2. Ober, C., Weitkamp, L. R., Cox, N., Dytch, H., Kostyu, D. & Elias, S. HLA and mate choice in humans. *Am. J. Hum. Genet.* **61,** 497–504 (1997).
3. Temple, S. D., Waples, R. K. & Browning, S. R. Modeling recent positive selection using identity-by-descent segments. *Am. J. Hum. Genet.* **111,** 2510–2529 (2024).
4. Albrechtsen, A., Moltke, I. & Nielsen, R. Natural selection and the distribution of identity-by-descent in the human genome. *Genetics* **186,** 295–308 (2010).
5. Kamiya, T., O'Dwyer, K., Westerdahl, H., Senior, A. & Nakagawa, S. A quantitative review of MHC-based mating preference: the role of diversity and dissimilarity. *Mol. Ecol.* **23,** 5151–5163 (2014).
6. Tregenza, T. & Wedell, N. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol. Ecol.* **9,** 1013–1027 (2000).
7. Wedekind, C. & Füri, S. Body odour preferences in men and women: do they aim for specific MHC combinations or simply heterozygosity? *Proc. Biol. Sci.* **264,** 1471–1479 (1997).
8. Havlíček, J., Winternitz, J. & Roberts, S. C. Major histocompatibility complex-associated odour preferences and human mate choice: near and far horizons. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **375,** 20190260 (2020).
9. Chaix, R., Cao, C. & Donnelly, P. Is mate choice in humans MHC-dependent? *PLoS Genet.* **4,** e1000184 (2008).
10. Scelza, B. A. Female choice and extra-pair paternity in a traditional human population. *Biol. Lett.* **7,** 889–891 (2011).
11. Scelza, B. A. & Prall, S. P. Only death will separate us: The role of extramarital partnerships among Himba pastoralists. *Arch. Sex. Behav.* **52,** 1355–1363 (2023).
12. Scelza, B. A., Prall, S. P., Swinford, N., Gopalan, S., Atkinson, E. G., McElreath, R., Sheehama, J. & Henn, B. M. High rate of extrapair paternity in a human population demonstrates diversity in human reproductive strategies. *Sci. Adv.* **6,** eaay6195 (2020).
13. Nemat-Gorgani, N., Guethlein, L. A., Henn, B. M., Norberg, S. J., Chiaroni, J., Sikora, M., Quintana-Murci, L., Mountain, J. L., Norman, P. J. & Parham, P. Diversity of KIR, HLA class I, and their interactions in seven populations of sub-Saharan Africans. *J. Immunol.* **202,** 2636–2647 (2019).
14. Swinford, N. A., Prall, S. P., Gopalan, S., Williams, C. M., Sheehama, J., Scelza, B. A. & Henn, B. M. Increased homozygosity due to endogamy results in fitness consequences in a human population. *Proc. Natl. Acad. Sci. U. S. A.* **120,** e2309552120 (2023).
15. Chen, J. J., Hollenbach, J. A., Trachtenberg, E. A., Just, J. J., Carrington, M., Rønningen, K. S., Begovich, A., King, M. C., McWeeney, S., Mack, S. J., Erlich, H. A. & Thomson, G. Hardy-Weinberg testing for HLA class II (DRB1, DQA1, DQB1, and DPB1) loci in 26 human ethnic groups: Hardy-Weinberg testing of HLA class II data. *Tissue Antigens* **54,** 533–542 (1999).
16. McQuillan, R., Eklund, N., Pirastu, N., Kuningas, M., McEvoy, B. P., Esko, T., Corre, T., Davies, G., Kaakinen, M., Lyytikäinen, L.-P., Kristiansson, K., Havulinna, A. S., Gögele, M., Vitart, V., Tenesa, A., Aulchenko, Y., Hayward, C., Johansson, A., Boban, M., Ulivi, S., Robino, A., Boraska, V., Igl, W., Wild, S. H., Zgaga, L., Amin, N., Theodoratou, E., Polašek, O., Girotto, G., Lopez, L. M., Sala, C., Lahti, J., Laatikainen, T., Prokopenko, I., Kals, M., Viikari, J., Yang, J., Pouta, A., Estrada, K., Hofman, A., Freimer, N., Martin, N. G., Kähönen, M., Milani, L., Heliövaara, M., Vartiainen, E., Räikkönen, K., Masciullo, C., Starr, J. M., Hicks, A. A., Esposito, L., Kolčić, I., Farrington, S. M., Oostra, B., Zemunik, T., Campbell, H., Kirin, M., Pehlic, M., Faletra, F., Porteous, D., Pistis, G., Widén, E., Salomaa, V., Koskinen, S., Fischer, K., Lehtimäki, T., Heath, A., McCarthy, M. I., Rivadeneira, F., Montgomery, G. W., Tiemeier, H., Hartikainen, A.-L., Madden, P. A. F., d'Adamo, P., Hastie, N. D., Gyllensten, U., Wright, A. F., van Duijn, C. M., Dunlop, M., Rudan, I., Gasparini, P., Pramstaller, P. P., Deary, I. J., Toniolo, D., Eriksson, J. G., Jula, A., Raitakari, O. T., Metspalu, A., Perola, M., Järvelin, M.-R., Uitterlinden, A., Visscher, P. M., Wilson, J. F. & ROHgen Consortium. Evidence of inbreeding depression on human height. *PLoS Genet.* **8,** e1002655 (2012).
17. Ceballos, F. C., Hazelhurst, S., Clark, D. W., Agongo, G., Asiki, G., Boua, P. R., Xavier Gómez-Olivé, F., Mashinya, F., Norris, S., Wilson, J. F. & Ramsay, M. Autozygosity influences cardiometabolic

disease-associated traits in the AWI-Gen sub-Saharan African study. *Nat. Commun.* **11,** 5754 (2020).
18. Inman, E. N. 'there are no true himbas anymore': Exploring the dynamics of the Himba culture and land use in the face of change in Kunene region, Namibia. *Sustainability* **16,** 1582 (2024).
19. Penn, D. J., Damjanovich, K. & Potts, W. K. MHC heterozygosity confers a selective advantage against multiple-strain infections. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 11260–11264 (2002).
20. Pierini, F. & Lenz, T. L. Divergent allele advantage at human MHC genes: Signatures of past and ongoing selection. *Mol. Biol. Evol.* **35,** 2145–2158 (2018).
21. Hazel, A., Marino, S. & Simon, C. An anthropologically based model of the impact of asymptomatic cases on the spread of Neisseria gonorrhoeae. *J. R. Soc. Interface* **12,** 20150067 (2015).
22. Hazel, A., Foxman, B. & Low, B. S. Herpes simplex virus type 2 among mobile pastoralists in northwestern Namibia. *Ann. Hum. Biol.* **42,** 543–551 (2015).
23. Hazel, A. & Holland Jones, J. Remoteness influences access to sexual partners and drives patterns of viral sexually transmitted infection prevalence among nomadic pastoralists. *PLoS One* **13,** e0191168 (2018).
24. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48,** W449–W454 (2020).
25. Nilsson, J. B., Kaabinejadian, S., Yari, H., Kester, M. G. D., van Balen, P., Hildebrand, W. H. & Nielsen, M. Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. *Sci. Adv.* **9,** eadj6367 (2023).
26. Charron, D. J., Lotteau, V. & Turmel, P. Hybrid HLA-DC antigens provide molecular evidence for gene trans-complementation. *Nature* **312,** 157–159 (1984).
27. Tollefsen, S., Hotta, K., Chen, X., Simonsen, B., Swaminathan, K., Mathews, I. I., Sollid, L. M. & Kim, C.-Y. Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. *J. Biol. Chem.* **287,** 13611–13619 (2012).
28. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11,** 823–836 (2011).
29. Meyer, D., C Aguiar, V. R., Bitarello, B. D., C Brandt, D. Y. & Nunes, K. A genomic perspective on HLA evolution. *Immunogenetics* **70,** 5–27 (2018).
30. Marnetto, D. & Huerta-Sánchez, E. *Haplostrips*: revealing population structure through haplotype visualization. *Methods Ecol. Evol.* **8,** 1389–1392 (2017).
31. Woelfing, B., Traulsen, A., Milinski, M. & Boehm, T. Does intra-individual major histocompatibility complex diversity keep a golden mean? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364,** 117–128 (2009).
32. Nowak, M. A., Tarczy-Hornoch, K. & Austyn, J. M. The optimal number of major histocompatibility complex molecules in an individual. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 10896–10899 (1992).
33. Radwan, J., Babik, W., Kaufman, J., Lenz, T. L. & Winternitz, J. Advances in the evolutionary understanding of MHC polymorphism. *Trends Genet.* **36,** 298–311 (2020).
34. Wegner, K. M., Kalbe, M., Kurtz, J., Reusch, T. B. H. & Milinski, M. Parasite selection for immunogenetic optimality. *Science* **301,** 1343 (2003).
35. Prall, S. P. & Scelza, B. A. Resource demands reduce partner discrimination in Himba women. *Evol. Hum. Sci.* **2,** e45 (2020).
36. Scelza, B. A. & Prall, S. P. Partner preferences in the context of concurrency: What Himba want in formal and informal partners. *Evol. Hum. Behav.* **39,** 212–219 (2018).
37. Prall, S. & Scelza, B. The effect of mating market dynamics on partner preference and relationship quality among Himba pastoralists. *Sci. Adv.* **8,** eabm5629 (2022).
38. Scelza, B., Prall, S. & Starkweather, K. The role of spousal separation on norms related to gender and sexuality among Himba pastoralists. *Soc. Sci. (Basel)* **10,** 174 (2021).
39. Kennedy, A. E., Ozbek, U. & Dorak, M. T. What has GWAS done for HLA and disease associations? *Int. J. Immunogenet.* **44,** 195–211 (2017).
40. Norman, P. J., Hollenbach, J. A., Nemat-Gorgani, N., Marin, W. M., Norberg, S. J., Ashouri, E., Jayaraman, J., Wroblewski, E. E., Trowsdale, J., Rajalingam, R., Oksenberg, J. R., Chiaroni, J., Guethlein, L. A., Traherne, J. A., Ronaghi, M. & Parham, P. Defining KIR and HLA class I genotypes at highest resolution via high-throughput sequencing. *Am. J. Hum. Genet.* **99,** 375–391 (2016).
41. Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M. & Chen, W.-M. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26,** 2867–2873 (2010).

42. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J. & Sham, P. C. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81,** 559–575 (2007).
43. Shkuri, O., Israeli, S., Tshuva, Y., Maiers, M. & Louzoun, Y. Efficient test for deviation from Hardy-Weinberg equilibrium with known or ambiguous typing in highly polymorphic loci. *Brief. Bioinform.* **25,** bbae416 (2024).
44. Delaneau, O., Zagury, J.-F. & Marchini, J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat. Methods* **10,** 5–6 (2013).
45. Spence, J. P. & Song, Y. S. Inference and analysis of population-specific fine-scale recombination maps across 26 diverse human populations. *Sci. Adv.* **5,** eaaw9206 (2019).
46. 1000 Genomes Project Consortium, Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., Korbel, J. O., Marchini, J. L., McCarthy, S., McVean, G. A. & Abecasis, G. R. A global reference for human genetic variation. *Nature* **526,** 68–74 (2015).
47. Freyman, W. A., McManus, K. F., Shringarpure, S. S., Jewett, E. M., Bryc, K., 23 and Me Research Team & Auton, A. Fast and robust identity-by-descent inference with the templated positional Burrows-Wheeler transform. *Mol. Biol. Evol.* **38,** 2131–2151 (2021).
48. Gusev, A., Lowe, J. K., Stoffel, M., Daly, M. J., Altshuler, D., Breslow, J. L., Friedman, J. M. & Pe'er, I. Whole population, genome-wide mapping of hidden relatedness. *Genome Res.* **19,** 318–326 (2009).
49. Nait Saada, J., Kalantzis, G., Shyr, D., Cooper, F., Robinson, M., Gusev, A. & Palamara, P. F. Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nat. Commun.* **11,** 6130 (2020).
50. Williams, C. M., Scelza, B. A., Daya, M., Lange, E. M., Gignoux, C. R. & Henn, B. M. A rapid, accurate approach to inferring pedigrees in endogamous populations. *bioRxiv* 2020.02.25.965376 (2020). doi:10.1101/2020.02.25.965376
51. Berger, S. A. GIDEON: a comprehensive Web-based resource for geographic medicine. *Int. J. Health Geogr.* **4,** 10 (2005).
52. Hazel, M.-A. Sexually Transmitted Diseases Among Pastoralists in Kaokoland, Namibia: Epidemiology, Ecology, and Behavior. (2012). at <https://scholar.google.com/citations?view_op=view_citation&hl=en&citation_for_view=QchpnBsAAAAJ:W7OEmFMy1HYC>
53. Zheng, X. & Lee, J. Imputation-based HLA typing with GWAS SNPs. *Methods Mol. Biol.* **2809,** 127–143 (2024).
54. Thomson, G. & Single, R. Conditional asymmetric linkage disequilibrium (ALD): Extending the biallelic r2 measure. *Genetics* **198,** 321–331 (2014).
55. Scelza, B. A., Atkinson, E. G., Prall, S., McElreath, R., Sheehama, J. & Henn, B. M. The ethics and logistics of field-based genetic paternity studies. *Evol. Hum. Sci.* **2,** 1–36 (2020).