Supplementary Materials

Supplementary Figures and Tables

Genomic evidence for MHC disassortative mating in humans

Claire Dandine-Roulland, Romain Laurent, Irene Dall'Ara, Bruno Toupance, Raphaëlle Chaix

Eco-Anthropologie, UMR 7206, CNRS, MNHN, Université Paris Diderot, Sorbonne Paris Cité, Paris, France

Supplementary Note

We performed a post-hoc power study to evaluate the influence of sample size on our ability to detect MHC non random mating signals from genomic data. To do so, we randomly extracted 1,000 subsamples of various sample size (from n = 36 to n = 156 couples) from the Netherlands population sample (n = 302 couples).

For each sample size, we estimated the power to detect MHC non random mating from genomic data, defined as the proportion of subsamples with two sided p-value below 0.05 (the p-value being computed as the proportion of permuted couples with more extreme relatedness coefficients than the real couples).
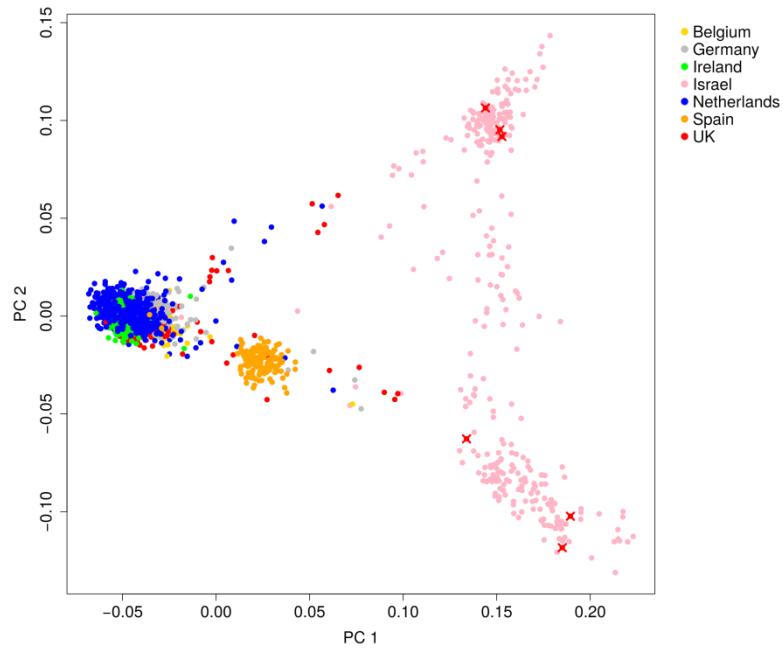
Figure S1: First two Principal Components of GAIN ADHD data. Crosses represent individuals sampled in UK matching with Israel samples and not included in the analysis.
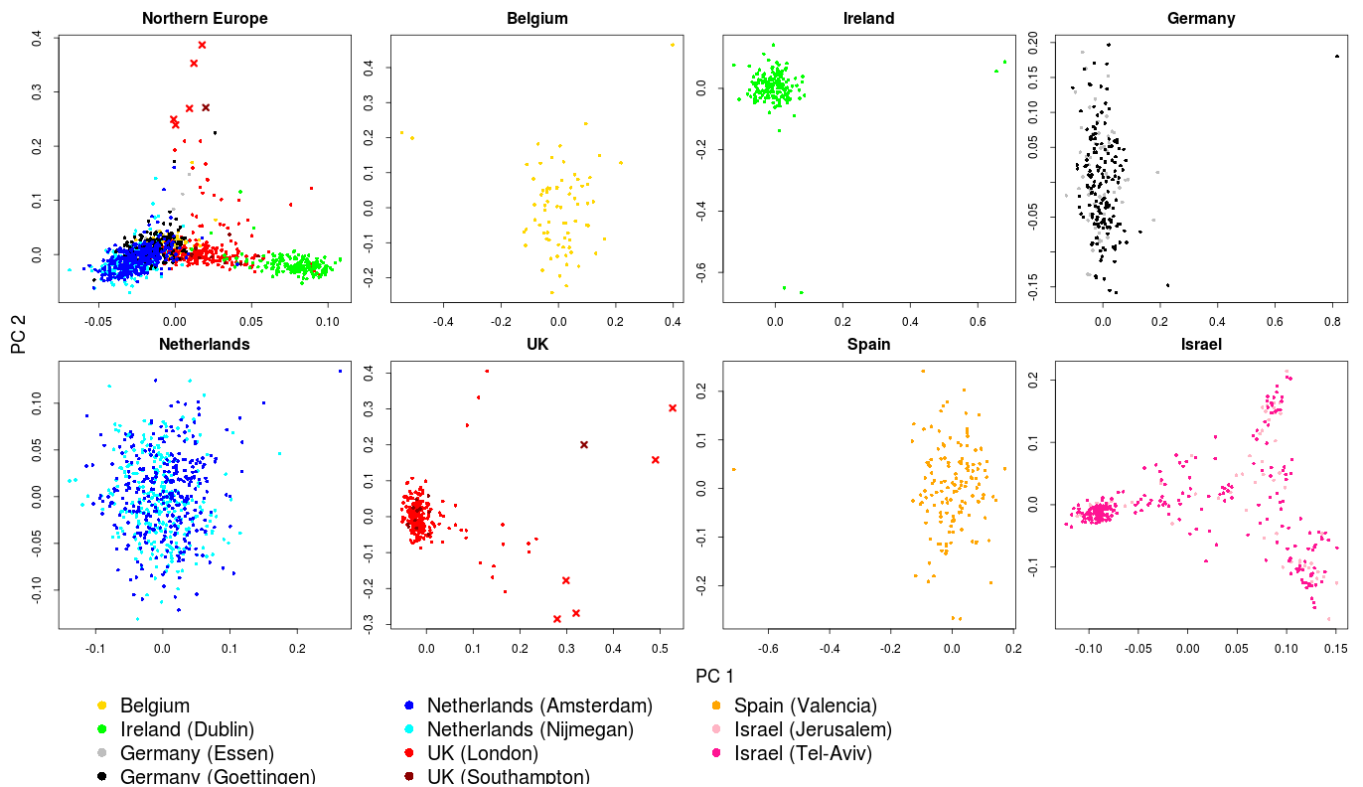

Figure S2: First two Principal Components of GAIN ADHD data by country and in the merged Northern European sample. Crosses represent individuals sampled in the UK matching with Israel samples and not included in the analysis.
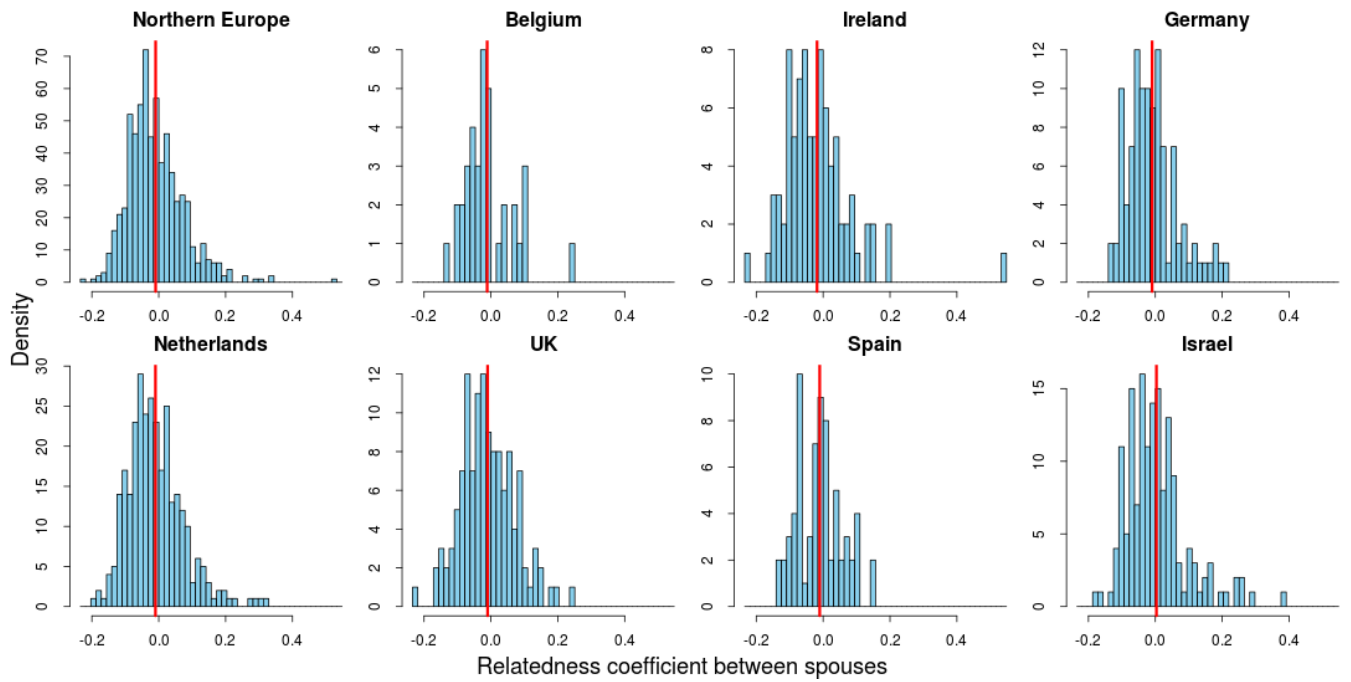
Figure S3: Distributions of MHC Rousset relatedness coefficients between spouses in each population. The red line corresponds to the mean relatedness coefficient between spouses.
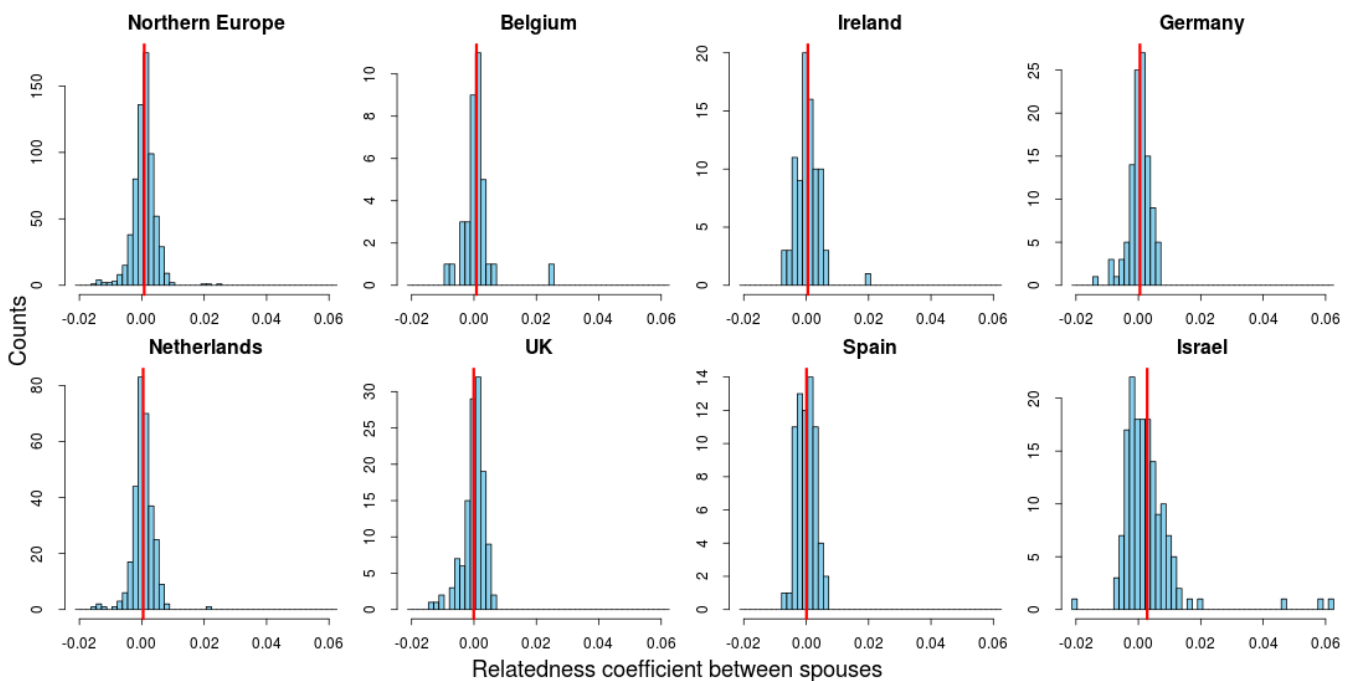

Figure S4: Distributions of genome-wide Rousset relatedness coefficients between spouses in each population. The red line corresponds to the mean relatedness coefficient between spouses.
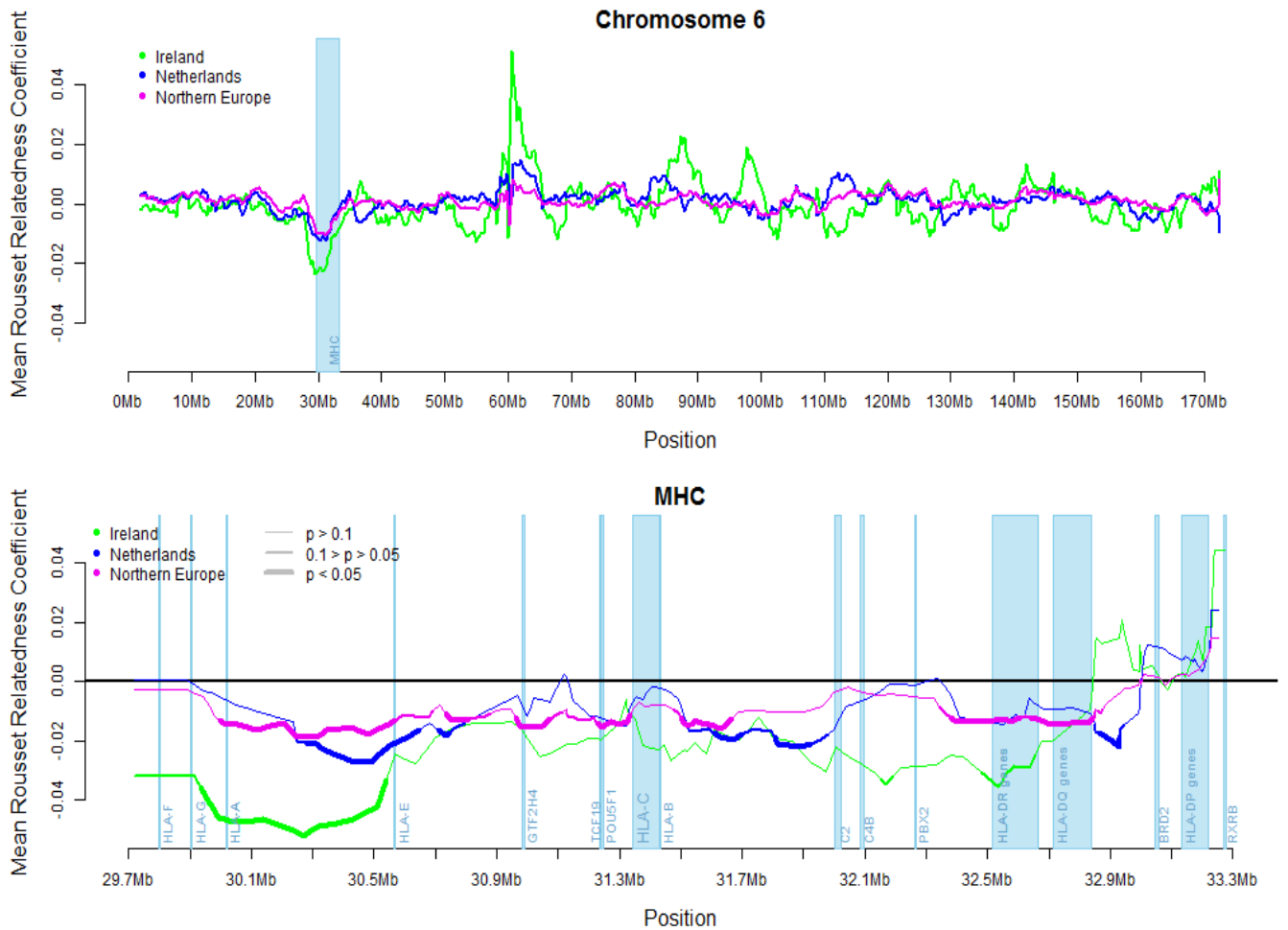
Figure S5: Mean Rousset relatedness coefficient between spouses along the MHC for sliding windows of 30 SNPs (in increments of 4 SNPs) in Northern Europe, Ireland and Netherlands. The width of lines represents the significance level of the non-random mating signal as assessed using a permutation test: for each window, 10^5 couples permutations were performed (attributing a new wife to each husband) and the p-value was defined as the proportion of permutations yielding a more extreme relatedness coefficient than the one observed for real spouses. Blue areas correspond to the location of the main genes within the MHC region.

Figure S6: Genome-wide Rousset relatedness versus MHC relatedness between spouses in Northern Europe, Ireland and Netherlands samples. The red point corresponds to the mean Rousset relatedness coefficient between spouses. The line corresponds to the linear regression line and the dotted lines to its 95% confidence interval.
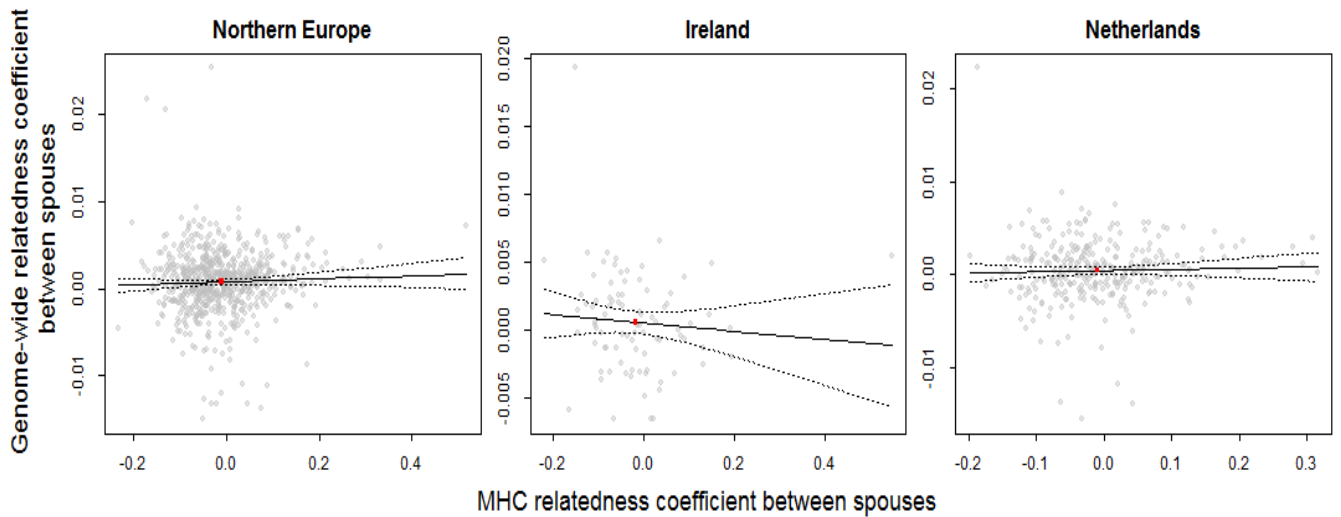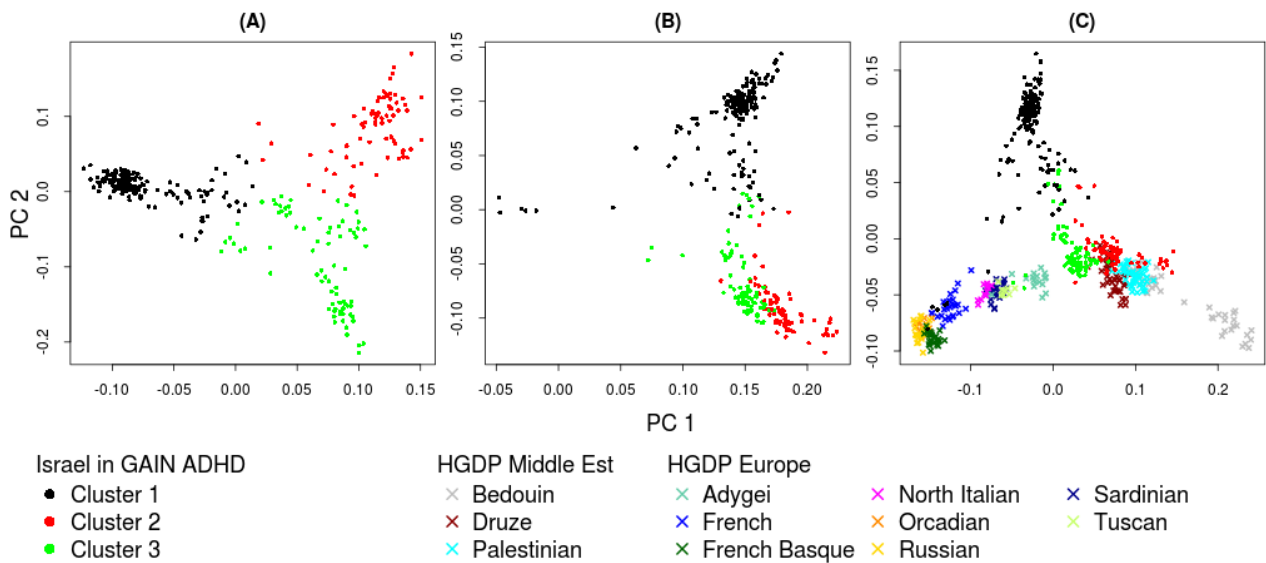

Figure S7: PCA analysis of the GAIN ADHD Israel samples merged with the HGDP European and Middle Eastern populations. A) PCA analysis of the GAIN ADHD Israel samples. We used the k-means clustering method to partition the Israel samples into three homogeneous clusters, shown here in black, green and red. B) GAIN ADHD Israel samples on PC1 and 2 when the PCA is performed on all GAIN ADHD samples. This figure is shown to allow identification of the three Israel clusters on the figure 1 of the main article. C) PCA analysis of GAIN ADHD Israel samples with the HGDP European and Middle Eastern samples.

Supplementary Tables

	MHC Heterozygosity
Northern Europe	0.326
Belgium	0.333
Ireland	0.325
Germany	0.330
Netherlands	0.330
UK	0.329
Spain	0.328
Israel	0.319

Table S1: Mean heterozygosity computed over the 421 MHC SNPs for each population sample

	Genome-wide		MHC		
Population	Rousset Relatedness Coefficient (R)	P-value[A]	Rousset Relatedness Coefficient (R)	P-value[B]	P-value[C]
Northern Europe	0.001	**< 10^{-5}**	-0.013	**0.008**	**9 x 10^{-4}**
Belgium	0.001	**0.029**	-0.018	0.381	0.217
Ireland	7 x 10^{-4}	0.058	-0.021	0.126	**0.005**
Germany	7 x 10^{-4}	**0.016**	-0.015	0.212	0.120
Netherlands	7 x 10^{-4}	**< 10^{-5}**	-0.014	**0.049**	**0.020**
UK	10^{-4}	0.744	-0.013	0.223	0.056
Spain	10^{-4}	0.867	-0.012	0.301	0.096
Israel	0.005	**< 10^{-5}**	0.005	0.704	0.559

Table S2: Mean Rousset relatedness coefficient (R) between spouses at the genome-wide level and at the MHC when setting to 1 (rather than 0.5) the proportion of variants identical by state between two heterozygotes. Bold values correspond to significant p-values considering the 5% threshold.
[A]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[B]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[C]two-sided p-value defined as the proportion of 3.6Mb genomic windows (with similar recombination rate as the MHC) exhibiting a more extreme R between spouses than the MHC

Population	Genome-wide		MHC		
	Rousset Relatedness Coefficient (R)	P-value[A]	Rousset Relatedness Coefficient (R)	P-value[B]	P-value[C]
Northern Europe	8×10^{-4}	**$< 10^{-5}$**	-0.021	**0.046**	**0**
Belgium	8×10^{-4}	0.300	-0.033	0.119	**0.019**
Ireland	5×10^{-4}	0.104	-0.031	0.324	**9×10^{-4}**
Germany	5×10^{-4}	0.321	-0.021	0.211	**0.010**
Netherlands	5×10^{-4}	**0.048**	-0.023	0.119	**0**
UK	-4×10^{-5}	0.112	-0.026	0.109	**0.003**
Spain	10^{-4}	0.998	-0.016	0.346	0.107
Israel	0.003	**$< 10^{-5}$**	-0.008	0.525	0.269

Table S3: Median of Rousset relatedness coefficient (R) between spouses at the genome-wide level and at the MHC. Bold values correspond to significant p-values considering the 5% threshold.
[A]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[B]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[C]two-sided p-value defined as the proportion of 3.6Mb genomic windows (with similar recombination rate as the MHC) exhibiting a more extreme R between spouses than the MHC

Population	Genome-wide		MHC		
	Genetic Correlation (C)	P-value[A]	Genetic Correlation (C)	P-value[B]	P-value[c]
Northern Europe	8×10^{-4}	**$< 10^{-5}$**	-0.017	**0.006**	**0**
Belgium	-0.012	**0.016**	-0.033	0.393	0.167
Ireland	-0.005	**0.029**	-0.038	0.066	**0.002**
Germany	-0.004	**0.007**	-0.022	0.256	0.064
Netherlands	-7×10^{-4}	**$< 10^{-5}$**	-0.018	0.066	**0.011**
UK	-0.004	0.894	-0.022	0.178	0.065
Spain	-0.007	0.696	-0.024	0.244	0.100
Israel	0.002	**$< 10^{-5}$**	0.004	0.594	0.762

Table S4: Mean genetic correlation (C) between spouses computed on standardized genotypes at the genome-wide level and at the MHC. Bold values correspond to significant p-values considering the 5% threshold.
[A]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme C than the C observed between real spouses
[B]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme C than the C observed between real spouses
[C]two-sided p-value defined as the proportion of 3.6Mb genomic windows (with similar recombination rate as the MHC) exhibiting a more extreme C between spouses than the MHC

Population	Genome-wide		MHC		
	Z-score	P-value[A]	Z-score	P-value[B]	P-value[C]
Northern Europe	0.244	**< 10⁻⁵**	-0.105	**0.003**	**0**
Belgium	0.252	**0.030**	-0.146	0.346	0.276
Ireland	0.187	**0.033**	-0.175	0.068	**0.045**
Germany	0.154	**0.009**	-0.119	0.171	0.095
Netherlands	0.144	**10⁻⁴**	-0.111	**0.036**	**0.030**
UK	-0.019	0.705	-0.111	0.168	0.101
Spain	0.038	0.727	-0.146	0.170	0.084
Israel	0.534	**< 10⁻⁵**	0.021	0.739	0.673

Table S5: Z-score indicating the deviation of Rousset relatedness coefficient between spouses and non-spouses at the genome-wide level and at the MHC. Bold values correspond to significant p-values considering the 5% threshold.
[A]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme Z-score than the Z-score observed for real spouses
[B]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme Z-score than the Z-score observed for real spouses
[C]two-sided p-value defined as the proportion of 3.6Mb genomic windows (with similar recombination rate as the MHC) exhibiting a more extreme Z-score than the MHC

Number of couples	Power
36	0.055
69	0.091
86	0.120
108	0.129
126	0.173
156	0.203

Table S6: Post Hoc power analysis: We randomly extracted 1,000 subsamples of various sample size (from n = 36 to n = 156 couples) from the Netherlands population sample (n = 302 couples). Here we show, for each sample size, the proportion of subsamples with two-sided p-value (proportion of permuted couples with more extreme relatedness coefficient than the real couples) below 0.05.

Sample	Correlation	P-value
Northern Europe	0.041	0.300
Ireland	-0.083	0.447
Netherlands	0.031	0.588

Table S7: Mean correlation between the Rousset relatedness coefficients computed at the MHC and at the genome-wide levels between Northern European, Irish and Dutch spouses, and corresponding p-values.

Population	Genome-wide		MHC		
	Rousset Relatedness Coefficient (R)	P-value[A]	Rousset Relatedness Coefficient (R)	P-value[B]	P-value[C]
Cluster 1 (black)	2×10^{-4}	0.595	2×10^{-4}	0.950	0.980
Cluster 2 (red)	0.006	**0.001**	0.036	0.100	0.070
Cluster 3 (green)	0.007	**10^{-4}**	9×10^{-4}	0.920	0.680

Table S8: Mean of Rousset relatedness coefficient (R) between spouses at the genome-wide level and at the MHC in the three Israel genetic clusters inferred from the PCA analysis. Bold values correspond to significant p-values considering the 5% threshold.
[A]two-sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[B]two sided p-value defined as the proportion of spouse permutations yielding a more extreme R than the R observed between real spouses
[C]two sided p-value defined as the proportion of 3.6Mb genomic windows (with similar recombination rate as the MHC) exhibiting a more extreme R between spouses than the MHC

Population	Relatedness between spouses (R)	Female mating preference (s)	Reduction of homozygosity (d)
Northern Europe	-0.009	0.0720 (h = 0) 0.0707 (h = 0.5) 0.0695 (h = 1)	0.0090
	-0.013[A]	0.0775 (h = 0) 0.0760 (h = 0.5) 0.0746 (h = 1)	0.0097
Netherlands	-0.010	0.0800 (h = 0) 0.0784 (h = 0.5) 0.0769 (h = 1)	0.010
	-0.014 [A]	0.0834 (h = 0) 0.0817 (h = 0.5) 0.0800 (h = 1)	0.0104

Table S9: Female mating preference (s) and reduction of homozygosity (d) inferred from mean Rousset relatedness (R) between spouses at the MHC using Hedrick model of female choice at equilibrium for 2 alleles and when the dominance parameter (h) for female preference is 0, 0.5 or 1.
[A] Mean Rousset relatedness coefficient (R) when setting to 1 (rather than 0.5) the proportion of variants identical by state between two heterozygotes.