European Journal of Human Genetics (2000) 8, 33–41

www.nature.com/ejhg

ARTICLE

A linkage disequilibrium map of the MHC region based on the analysis of 14 loci haplotypes in 50 French families

Alicia Sanchez-Mazas[1,2], Sami Djoulah[3], Marc Busson[3], Isabelle Le Monnier de Gouville[3], Jean-Claude Poirier[3], Catherine Dehay[3], Dominique Charron[3], Laurent Excoffier[1,2], Stefan Schneider[1,2], André Langaney[1,2], Jean Dausset[4] and Jacques Hors[3]

[1]Laboratory of Genetics and Biometry, Department of Anthropology and Ecology, University of Geneva, Switzerland; [2]Musée de l'Homme, Paris, France; [3]LNH, Department of Immunology and INSERM U396, Saint-Louis Hospital, Paris, France; [4]CEPH, Fondation Jean Dausset, Saint-Louis Hospital, Paris, France

A sample of 100 individuals from 50 French families of known pedigrees were typed for 14 loci of the HLA region (DPB1, DQB1, DQA1, DRB1, DRB3, 4, 5, C4B, C4A, Bf, C2, TNFa, TNFb, B, Cw, A). Linkage disequilibrium in each pair of loci was investigated by an exact test using a Markov chain algorithm. The results indicate no disequilibrium between DPB1 and the other loci, whereas the other class II genes are all significantly linked to each other. Linkage disequilibrium is also detected between some pairs of class I and class II-class I loci despite the long physical distance separating the loci (eg A–B, Cw–DRB1). On the other hand, some contiguous loci of the class III region are found to be in equilibrium with each other. Several hypotheses including selection, but also unequal allelic diversity at different MHC loci are discussed to explain this complex pattern of linkage disequilibrium. *European Journal of Human Genetics* (2000) **8**, 33–41.

Keywords: MHC; linkage disequilibrium; HLA haplotypes; French population; CEPH families; population genetics

Introduction

The human major histocompatibility complex (HLA) is located within a DNA region of about 4 Mb on chromosome 6, band 6p21.3. The detailed physical map of this chromosomal segment reveals a high number of HLA genes and pseudogenes (class I, class II and the non-classical class III), in addition to other loci of HLA-related, unrelated, or unknown functions.[1] Most of these genes exhibit allelic variation, some of them (eg *DRB1* and B loci) being among the most polymorphic so far known in human populations.[2]

The phenomenon of linkage disequilibrium has become a prominent characteristic of the MHC since the early descriptions of the multilocus polymorphism.[3,4] Since then, numerous 'extended' haplotypes have been described in human populations all around the world.[5,6] Their likely conservation in human populations has led some authors to define the notion of 'ancestral' haplotypes, preserved for various possible reasons from early human populations.[7,8] However, the choice of an accurate statistical measure of linkage disequilibrium, as well as its biological interpretation, is not straightforward.[9,10] One reason why linkage disequilibrium is often poorly estimated is that it requires an accurate estimation of haplotype frequencies, which is difficult when sample sizes are small.

Family data are most informative for defining the haplotypic profile, and hence for assessing the pattern of linkage disequilibrium in a given population. In this context, the data recorded by the Centre d'Etude du Polymorphisme Humain (CEPH) represent unique material, not only because of extended pedigree information, but also because they constitute extensive work and shared database. Previous reports using international CEPH family data have provided relevant information on linkage disequilibrium and recombinant haplotypes across the HLA region, mostly the class II region.[11–17] In the present study, the data related to 14 HLA

Correspondence: A Sanchez-Mazas, Laboratory of Genetics and Biometry, Department of Anthropology and Ecology, University of Geneva, 12 rue Gustave-Revilliod, CH-1227 Carouge, Geneva, Switzerland. Tel: +41 22 702 69 84; Fax: +41 22 300 03 51; E-mail: sanchez@sc2a.unige.ch.
Received 3 February 1999; revised 23 June 1999; accepted 3 August 1999

loci distributed across the whole MHC region in 50 CEPH families are used. These families were especially chosen for their known French origin, and only 10 of them appear in previously published data. Original HLA typings, including Cw DNA sequencing and the non-classical class III, have been undertaken here on all families. The results are applied to an original investigation testing the global linkage disequilibrium of each pair of loci under study (91 pairs for 14 loci) using a recently developed non-parametric statistical approach (Markov chain). Our aim is to obtain a reliable statistical assessment of the pattern of global linkage disequilibrium among the loci of the MHC, in addition to detection of specific allelic associations tested by more conventional methods. This should allow the location of possible recombination hotspots or coldspots in the MHC region to be identified, and extended HLA linkage groups in this population to be defined.

Subjects and methods

Subjects

The sample consists of 100 independent individuals (parents) from 50 French families. Appropriate informed consent was obtained from all individuals. Families were recruited from the Paris area, with familial origin (parents and grandparents) in all parts of France. The children of each family represent 275 offspring with 4–10 children per family.

HLA typing

A total of 14 loci located within the 4 Mb MHC region was tested on the 100 parents, and either all or a subset of these loci on the 275 offspring. The alleles of the loci DRB3, DRB4, and DRB5, which are mutually exclusive, were considered here to belong to a unique locus. The resulting 14 loci represent three class I (B, Cw, A), five class II (DPB1, DQB1, DQA1, DRB1, DRB3, 4, 5), and six class III (C4B, C4A, Bf, C2, TNFa, TNFb) polymorphisms.

Class I (A, B) specificities were tested by a microlymphocytoxicity technique adapted from Mittal *et al*[18] to the lymphoblastoid cell lines as target of alloreactive specific human sera.[19] A total of 110 mono or bispecific allosera was used with a correlation *r* value for specificity assignment of 0.9–1, for locus A, and 0.8–1, for locus B. HLA-Cw alleles were studied by direct PCR exon 2–3 sequencing, using ALF express™, with a Cw HLA typing kit (Pharmacia Biotech, Saclay, France).

Class II polymorphisms were defined using PCR/SSO reverse hybridisation technique (Innolipa, Louvain-La-Neuve, Belgium), for HLA-DRB1, DRB3, 4, 5, DQB1 and DPB1 alleles, and PCR/SSO, for DQA1 alleles. For DRB1 and DRB3, 4, 5, direct sequencing analysis using ALF express™ DNA sequencer was also performed, where the second exon was amplified by PCR DRB group-specific primers with a DRB HLA typing kit (Pharmacia Biotech).[20]

Class III complement components were tested, for C2, by electrophoresis pattern on polyacrylamid gel[21] and, for C4A and C4B, with an additional test of functional assay.[22] The Bf polymorphism was studied following Alper *et al*.[23] The tumour necrosis factor (TNF) region was investigated by testing two microsatellite polymorphisms, TNFa and TNFb.[24,25]

Statistical tests

The gametic phase of the 100 parental genotypes could be established on the basis of the known pedigrees, allowing for estimation of allele and haplotype frequencies by direct gene counting. Standard deviations of allele frequencies were computed from binomial variances. Hardy-Weinberg equilibrium was tested at each locus using an exact test based on a Markov chain of 100 000 steps.[26] The same test was also applied to the loci pairs in order to detect a possible relationship with linkage disequilibrium of two-locus haplotypes.

The significance of the linkage disequilibrium between each pair of loci was estimated using an exact test based on a Markov chain approach.[27,28] The method enables the exploration, under the null hypothesis of no linkage disequilibrium, of the probability distribution of the contingency tables with the same marginal totals as the observed one. The *P*-value is the proportion of tables with a lower probability than the observed one. This test aims to detect a global linkage disequilibrium between two given multiallelic loci (non-independent segregation), rather than a specific linkage disequilibrium between two particular alleles of two loci (allelic association). The distinction between these two approaches is explained by Slatkin.[28] The Markov chain was run in this case for one million stages. In addition, the classic coefficients of linkage disequilibrium, D (and its significance using a χ^2 test), and of normalised linkage disequilibrium, D', between the alleles of two given loci, were computed for each individual haplotype according to the methods described by Lewontin[10] and Weir.[29]

A test of selective neutrality[30,31] was performed for each locus except TNF. This test compares the observed and expected values of the F statistic (equal to the sum of squared allele frequencies), and detects any departure from neutral expectation, assuming population equilibrium. The empirical distribution of F was generated by simulating a series of 1000 random samples with a number of alleles and a sample size identical to those of the observed data. The test was not performed for the TNF loci, because the infinite allele model conditioning this test[32] does not apply to microsatellite polymorphisms.

All statistical tests were carried out with the program package ARLEQUIN.[33]

Results

Allele frequencies

Detailed allele frequencies at each locus are available on the Internet at http://anthropologie.unige.ch/~alice/ejhg99/. No deviation from Hardy-Weinberg equilibrium is observed for any locus (1% level). Class I and class II (with the exception of DPB1) and TNF (a and b) allele distributions exhibit many low frequency alleles, whereas DPB1 and the other class III loci allele distributions are uneven due to the presence of one very frequent allele. The observed frequencies are generally in agreement with previously published frequencies in European populations,[5,6,34,35] with frequent A2 (24%), A3 (13%), A1 (10.5%), B44 (15%), B35 (7%), B8 (11%), B18 (10%, this allele being frequent in Greeks), Cw*0701 (14.5%), Cw*0501 (11%), DRB1*0701 (13%), DRB1*0301 (12.5%), DRB1*0101 (9.5%), DQA1*0101 (17.5%), DQA1*0102 (15%), DQA1*0501 (26.5%, this allele being very frequent in the South-East French[34]), DQB1*0201 (25%), DQB1*0301 (19.5%), DQB1*0501 (11.5%), DPB1*0401 (39%), BfS (74.5%), C4A3 (78%), and C4B1 (71.5%). For the class II loci, the present data differ from the CEPH data previously analysed by Begovitch *et al*[11] by higher frequencies for DQB1*0201, DRB1*0301, DRB1*0101, and lower frequencies for DQB1*0302, DQB1*0602, DRB1*1501 in the former. These may be due to differences in the composition of the studied samples (French families *vs* 'international CEPH families', respectively).

The TNFa and TNFb polymorphisms were previously shown to be highly heterogeneous among European populations.[36,37] Higher frequencies of TNFa6 (18%), a10 (18.5%) and b5 (36.5%), and lower frequencies of TNFa11 (4.5%) and b6 (2%) are here found compared with the French data analysed by Crouau-Roy *et al*.[36] Distinct origins of the individuals studied in each sample may account for these differences.

Haplotype frequencies and linkage disequilibrium on individual haplotypes

Only one extended haplotype is observed twice in the sample (HLA DPB1*0402–DQB1*0501–DQA1*0101–DRB1*0101–C4B*Q0 – C4A*2 – Bf*F – C2*1 – TNFa*5 – TNFb*5– B35–Cw*0401–A3), giving a total number of 199 different 14-locus haplotypes among 200 chromosomes. The raw data is available on request. The frequencies of class I, class II, and class I-class II haplotypes for the most commonly studied loci combinations are given in Table 1. The most frequent DQB1–DRB1–B–Cw–A haplotype observed (DQB1*0201–DRB1*0301–B8–Cw*0701–A1) has a frequency of only 2%. A considerable amount of polymorphism is thus observed on this French sample, and we do not detect any relevant class I-II-III extended haplotype.

Detailed frequencies and coefficients of linkage disequilibrium D and D' are given on the Internet at http://anthropologie.unige.ch/~alice/ejhg99/ for the most frequent two-locus haplotypes. It is noteworthy that the extended haplotype DQB1*0201–DRB1*0301–B8–Cw*0701–A1, usually reported as common in Europeans, presents significant allelic associations for each pair of adjacent loci (DQB1*0201–DRB1*0301, DRB1*0301–B8, B8–Cw*0701, and Cw*0701–A1). The same is true of the DQB1*0201–DRB1*0701–B44–Cw*1601–A29, also commonly found in Europeans. This may suggest that linkage disequilibrium extends over multiple loci in several haplotypes. However, whereas some class I (eg B8–Cw*0701–A1) and many class II (eg DQB1*0201–DQA1*0201–DRB1*0701–DRB4*01011/3) haplotypes present significant allelic associations between all possible pairs of loci (Table 1a and 1b) this result is never true of extended class II-class I haplotypes (Table 1c). Thus the hypothesis that extended class II-class I haplotypes may be conserved through time (and hence be considered as ancestral) still needs to be supported by a reliable measure of multipoint linkage disequilibrium.

Global linkage disequilibrium between HLA loci

Table 2 presents the significance of global linkage disequilibrium between each pair of HLA loci, as found by the Markov chain procedure. The chosen significance level ($\alpha = 0.01$) was corrected by the number of tests (91 tests for 14 loci) performed on the data (Bonferroni procedure[29]). The pairs of loci which were found in significant linkage disequilibrium are reported on the MHC chromosome map shown in Figure 1, allowing the observed pattern of linkage disequilibrium and the physical distance between the loci to be visualised simultaneously.

No significant disequilibrium is detected between DPB1 and any class I, II or III loci, whereas the other class II loci DQB1, DQA1, DRB1 and DRB3, 4, 5, which are physically close, are all found in significant disequilibrium with each other. Two additional sets of physically close loci are in significant disequilibrium: TNF (a and b), B and Cw, on the one hand, and C4A–C4B on the other. Significant linkage disequilibrium is also observed between some physically distant loci ('long distance' linkage disequilibrium): between the A and B class I loci, and between some class II and class I loci (DQA1–B, DRB1–Cw, and DRB3, 4, 5–Cw). Conversely, no linkage disequilibrium is observed between some physically very close loci, such as Bf–C2, Bf–C4 (A and B), and C2–C4 (A and B).

Selective neutrality tests

Table 3 presents the results of the Ewens-Watterson tests of selective neutrality carried out separately for each locus (except TNFa and TNFb). All HLA class I loci (A, B, Cw) exhibit a lower homozygosity than expected under the neutral hypothesis (5% unilateral test). This is also the case for DQA1. On the other hand, the hypothesis of neutrality is not rejected for DPB1, DQB1, DRB1, and DRB3, 4, 5, nor for the electrophoretical class III loci.

To investigate further a possible selection on the MHC loci, we tested the Hardy-Weinberg equilibrium on all possible loci

pairs. A deviation from equilibrium could involve multiple causes related either to the pairs of loci under study (eg selection on specific haplotypes) or to the population itself (eg migration or non-panmictic state). However, we did not find any significant rejection of the Hardy-Weinberg equilibrium except for the DQA1–DQB1 loci pair. We only note peculiar behaviour of locus DQA1 as, in addition to the rejection of the Hardy-Weinberg equilibrium for the pair DQA1–DRB1, it is the only class II locus for which the Ewens-Watterson test is also significant (Table 3).

Discussion

Overall, the present study reveals a very complex pattern of linkage disequilibrium throughout the MHC region in this particular French population. There are at least five relevant observations.

1 Significant linkage disequilibrium is found between all HLA class II loci except DPB1 In addition to the significant global linkage disequilibrium found between these loci

Table 1 HLA class I (1), class II (2), and class II-class I (3) most frequent haplotypes in 100 unrelated French parents

			Significant allelic associations[c]			
1 *Class I* *B–Cw–A*	*Freq.*[a] (>1.5%)	*s.d.*[b]	*B–Cw*	*B–A*	*Cw–A*	*Total*
8–0701–1	0.035	0.013	*	*	*	3
44–1601–29	0.030	0.012	*	*	*	3
44–0501–2	0.030	0.012	*	–	–	1
35–0401–3	0.020	0.010	*	*	*	3
55–0303–11	0.020	0.010	*	*	*	3
7–0702–3	0.020	0.010	*	–	–	1
8–0701–2	0.015	0.009	*	–	–	1
18–0501–30	0.015	0.009	*	–	–	1
18–1203–25	0.015	0.009	–	*	*	2

			Significant allelic associations						
2 *Class II* *DQB1–DQA1–DRB1–DRB3,4,5*	*Freq.* *(>3%)*	*s.d.*	*DQB1–DQA1*	*DQB1–DRB1*	*DQB1–DRB3,4,5*	*DQA1–DRB1*	*DQA1–DRB3,4,5*	*DRB1–DRB3,4,5*	*Total*
0201–0201–0701–B4*01011/3	0.115	0.023	*	*	*	*	*	*	6
0501–0101–0101–nogene	0.085	0.020	*	*	*	*	*	*	6
0602–0102–15011–B5*0101	0.075	0.019	*	*	*	*	*	*	6
0301–0501–1101–B3*0202	0.070	0.018	*	*	*	*	*	*	6
0201–0501–0301–B3*0101	0.065	0.017	*	*	*	*	*	*	6
0201–0501–0301–B3*0202	0.055	0.016	*	*	–	*	*	–	4
0503–0101–1401–B3*0202	0.045	0.015	*	*	*	*	–	*	5
0603–0103–1301–B3*0202	0.045	0.015	*	*	*	*	–	*	5
0302–0301–0401–B4*01011/3	0.040	0.014	*	*	*	*	*	*	6
0302–0301–0404–B4*01011/3	0.035	0.013	*	*	*	*	*	*	6
0604–0102–1302–B3*0301	0.035	0.013	*	*	*	*	*	*	6
0301–0501–1104–B3*0202	0.035	0.013	*	*	*	*	*	*	6
0301–0301–0401–B4*01011/3	0.030	0.012	–	*	–	*	*	*	4

			Significant allelic associations									
3 *Class II-Class I* *DQB1–DRB1–B–Cw–A*	*Freq.* *(>1%)*	*s.d.*	*DQB1–DRB1*	*DQB1–B*	*DQB1–Cw*	*DQB1–A*	*DRB1–B*	*DRB1–Cw*	*DRB1–A*	*B–Cw*	*B–A*	*Cw–A*	*Total*
0201–0301–8–0701–1	0.020	0.010	*	*	*	–	*	*	–	*	*	*	8
0201–0701–44–1601–29	0.015	0.009	*	–	–	–	*	–	*	*	*	*	6
0602–15011–7–0702–3	0.015	0.009	*	*	–	–	*	–	–	*	–	–	4
0501–0101–35–0401–3	0.015	0.009	*	*	*	–	–	–	–	*	*	*	6
0201–0301–8–0701–2	0.010	0.007	*	*	*	–	*	*	–	*	–	–	6
0301–0401–44–0501–2	0.010	0.007	*	–	–	–	–	–	–	*	–	–	2
0201–0301–8–0702–2	0.010	0.007	*	*	–	–	*	–	–	–	–	–	3
0503–1401–55–0303–11	0.010	0.007	*	*	–	*	*	–	*	*	*	*	8
0201–0301–18–0501–31	0.010	0.007	*	–	–	–	–	–	–	*	–	–	2
0201–0301–8–0701–26	0.010	0.007	*	*	*	–	*	*	–	*	–	–	6
0201–0701–13–0602–30	0.010	0.007	*	–	–	–	*	*	–	*	*	*	6
0301–1101–44–0501–2	0.010	0.007	*	–	–	–	–	–	–	*	–	–	2

[a]Freq. = allele frequency
[b]s.d. = standard deviation
[c]*: the allelic association is significant at the 1% level; –: not significant.

Table 2 Linkage disequilibrium significance level between each pair of MHC loci under study

Locus	*DPB1*	*DQB1*	*DQA1*	*DRB1*	*DRB3,4,5*	*C4B*	*C4A*	*Bf*	*C2*	*TNFBα*	*TNFBβ*	*B*	*Cw*	*A*
DPB1	*	0.3954	0.2246	0.1019	0.3741	0.3396	0.5471	0.2130	0.2584	0.0096	0.3075	0.1900	0.6794	0.2336
DQB1	–	*	0	0	0	0.4653	0.2058	0.0819	0.1843	0.0097	0.2832	0.0017	0.0139	0.6775
DQA1	–	+	*	0	0	0.0195	0.0757	0.0434	0.0456	0	0.1918	0	0.0004	0.5977
DRB1	–	+	+	*	0	0.0303	0.2759	0.2042	0.0389	0.0091	0.1095	0.0027	0	0.0292
DRB3,4,5	–	+	+	+	*	0.2469	0.0134	0.7170	0.6838	0.0004	0.0298	0.0006	0	0.8100
C4B	–	–	–	–	–	*	0	0.0050	0.0030	0.0044	0	0.0288	0.0018	0.1413
C4A	–	–	–	–	–	+	*	0.0519	0.0125	0.0470	0	0.0005	0	0.0127
Bf	–	–	–	–	–	–	–	*	0.1751	0.00007	0.2023	0	0.0255	0.2746
C2	–	–	–	–	–	–	–	–	*	0.0945	0.1129	0.0193	0.0081	0.2464
TNFBα	–	–	+	–	–	–	–	+	–	*	0	0	0	0.2664
TNFBβ	–	–	–	–	–	+	+	–	–	+	*	0	0	0.5878
B	–	–	+	–	–	–	–	+	–	+	+	*	0	0
Cw	–	–	–	+	+	–	+	–	–	+	+	+	*	0.0060
A	–	–	–	–	–	–	–	–	–	–	–	+	–	*

Note. Above diagonal: probability values obtained after 1 000 000 steps of the Markov chain procedure. Below diagonal: 1% significance after Bonferroni's correction for multiple tests. A significant value is indicated by +, and a non-significant value by –.

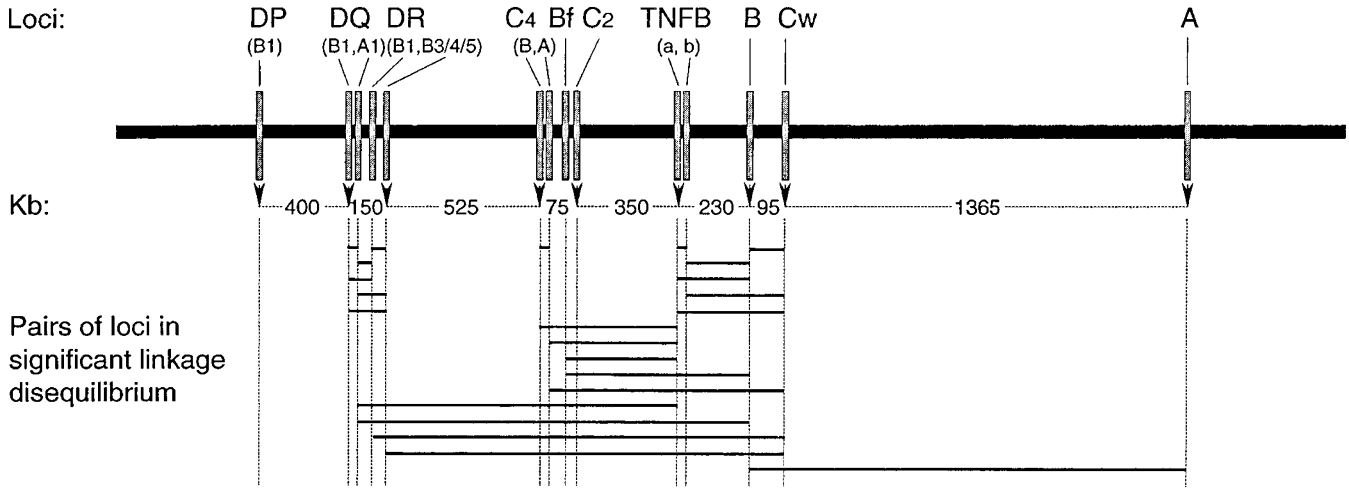

Figure 1 Map of the human MHC region showing the physical distances between the 14 loci under study. The horizontal bars plotted below the physical map show the pairs of loci in significant linkage disequilibrium (see Table 2 for details).

Table 3 Results of the Ewens-Watterson test of selective neutrality

Locus	*Observed homozygosity*	*Expected homozygosity*	*Lower limit of 10% confidence interval (5% unilateral test)*	*Significance*[a]
DPB1	0.1964	0.1483	0.0915	ns
DQB1	0.1439	0.2460	0.1413	ns
DQA1	0.1675	0.3407	0.1897	**
DRB1	0.0786	0.1043	0.0666	ns
DRB3,4,5	0.2097	0.3425	0.1892	ns
C4B	0.5394	0.5320	0.2917	ns
C4A	0.6258	0.4746	0.2582	ns
Bf	0.6063	0.6127	0.3509	ns
C2	0.8961	0.8275	0.5113	ns
B	0.0724	0.1189	0.0770	**
Cw	0.0821	0.1657	0.0951	**
A	0.1133	0.1995	0.1188	**

[a]ns: not significant; **: significant at the 5% level.

(Table 2), many individual DQB1–DQA1–DRB1–DRB3, 4, 5 haplotypes detected in the present sample exhibit both a high frequency (3–12%) and significant allelic associations between all pairs of loci (Table 1b). Both statistical approaches here applied to detect linkage disequilibrium (test for the significance of global linkage disequilibrium among loci, and test for the significance of allelic associations) agree with the hypothesis that the loci DQA1, DQB1, DRB1 and DRB3, 4, 5 form a tight linkage group.

This is in agreement with the conclusions of previous population studies, where a high level of linkage disequilibrium is usually detected for specific DQB1–DRB1 or DQB1–DQA1–DRB1 haplotypes.[6] Moreover, no recombination, except for one recently reported case,[38] has ever been observed between DRB1 and DQB1 (results confirmed by the analysis of the 12th HLA workshop recombinant families[39]). Close associations are also found between the DRB1 and DRB3, 4, 5 alleles (DRB1*03, 11, 12, 13, 14 and DRB3, DRB1*04, 07, 09 and DRB4, DRB1*15, 16 and DRB5, and DRB1*01, 08, 10 and the absence of a second *DRB* gene). Only a strong linkage disequilibrium between DRB1 and DRB3, 4, 5 could have maintained these associations after the likely emergence of the second *DRB* gene by duplication from DRB1.[40]

2 No linkage disequilibrium is detected between DPB1 and other HLA loci This result should be considered together with other previous studies. On the one hand, extended DP–DR and DP–DQ–DR haplotypes have been found in positive linkage disequilibrium in previously studied international CEPH families,[11,13] as well as in some specific human populations like the Cayapa in Ecuador.[41] On the other hand, class II linkage disequilibrium involving DPB1 alleles are found to be weak in several populations of European origin,[42] and many studies have suggested the presence of recombination hotspots in the region between the DPB1 and DQB1 loci. Total linkage equilibrium has been suggested between the genes encoding the transporters TAP1 and TAP2[14] and three preferential recombination sites have been proposed between HLA–DNA and RING3, between DQB3 and DQB1, and within TAP2.[15] Our studies may not conflict with previous observations. Allelic associations between some DPB1 and DQ or DR alleles may exist but may not be strong enough to create a significant global linkage disequilibrium between the corresponding loci. Alternatively, testing allelic associations on individual haplotypes by parametric methods, such as χ^2 tests, may give false positives (type I error) when sample sizes are small, as is commonly the case in population genetic studies.

3 The significance of linkage disequilibrium is not necessarily related to the physical distance between the loci Like the four loci DQB1, DQA1, DRB1 and DRB3, 4, 5, the loci B, Cw and TNF (a and b) appear to form a tight linkage group, where each locus is in significant linkage disequilibrium with the other. As for DQ–DR (about 150 kb between DQB1 and DRB3, 4, 5), the physical proximity of these loci (around 300 kb between TNF and Cw), compared with the whole MHC region, may explain this result, a negative correlation between linkage disequilibrium and physical distance being generally expected except for very short genomic regions.[43] However, the present study fails to demonstrate a clear-cut relationship between linkage disequilibrium and physical distance. For example, a significant linkage disequilibrium is detected for the A–B pair despite the very long physical distance separating locus A from the other class I loci (about 1500 kb). Such a result is unlikely to have occurred by chance alone, as it is in agreement with classic population genetic studies showing, for example, that the haplotype A1–B8 is conserved in Europe. It also supports the estimation of genetic distances between the HLA loci, as a very low recombination rate has been estimated for the class I region (0.31% between A and B, *vs* 0.74% between DPB1 and DRB1, and 0.94% between DRB1 and B[16]) compared with what one would expect from physical distance (an average of 1% recombination per megabase). Other cases of significant linkage disequilibrium between physically distant loci are found in this study (between DQA1 and B, between DRB1 and Cw, and between DRB3, 4, 5 and Cw), together with significant allelic associations (like B8–DRB1*0301 and B44–DRB1*0701, common in Europeans). As noted previously for anonymous regions of the genome,[44] the conservation of chromosome segments along the HLA region, possibly varying from one population to another,[8,45,46] is thus not a systematic consequence of physical linkage,[47] but may be due to other mechanisms, such as natural selection (see below).

4 Both directional and balancing selection may influence the evolution of MHC Natural selection is most often invoked to influence the evolution of MHC and may partly account for the linkage disequilibrium pattern observed in the HLA region. For example, a powerful protective effect would be conferred by haplotypes DRB1*1501–DQA1*0102–DQB1*0602, DRB1*1301–DQA1*0103–DQB1*0603, and DRB1*07–DQA1*0301–DQB1*0201 in case of insulin-dependent diabetes mellitus (IDDM).[48,49] The significant linkage disequilibrium here detected between some physically distant loci, like A and B, at both the locus and allelic levels, indicates that some form of directional selection (due, for example, to the existence of molecular cooperation between HLA molecules during the immune response, and/or to protective effects conferred by specific haplotypes) may influence the evolution of MHC over extended genomic regions. This would account for the conservation of some HLA haplotypes.

On the other hand, balancing selection may be responsible for the maintenance of a high level of polymorphism at several HLA loci. This was previously supported by a higher rate of replacement than silent mutations in the MHC region,[50–52] the continuance of very old HLA lineages (older than 5 million years) in present human populations,[53–55] and

a departure from neutral expectation of HLA distributions demonstrated by statistical tests.[56–58] The present study sustains the hypothesis of balancing selection acting on MHC, but indicates a possible difference between the class I and class II loci. Indeed, a significantly low homozygosity is observed for all class I loci, whilst the class II loci, except for DQA1, appear to be selectively neutral (Table 3). These results agree with previous neutrality tests performed on HLA[17,58,59] and the estimation of a very weak selective effect for DRB1.[57] To explain the high diversity of class I molecules, we may suggest that the presentation of a wider variety of endogenous peptides, for example, from viruses, by the HLA class I molecules of a heterozygous individual would give him a higher probability of survival against such pathogens. The effect would be weaker for class II molecules whose antigen binding site is wider and possibly more ubiquitous. Although a causal relationship between linkage disequilibrium and balancing selection is questionable at the present stage of research, both phenomena are here shown to occur simultaneously on HLA class I loci, whereas this is not the case for most HLA class II loci.

5 Amount of linkage disequilibrium observed is a function of allelic diversity The pattern of linkage disequilibrium observed in the present study for the class III loci presents some apparent inconsistencies. Indeed, no statistical association is found between some physically close loci like Bf and C2 (Table 2 and Figure 1), while there is no known indication of a recombination hotspot in this region. No significant linkage disequilibrium is found either between the class II and C4–Bf–C2 loci, although we showed that some class II loci were linked to the more distant TNF, B and Cw loci. One hypothesis is that the statistical test failed to detect linkage disequilibrium for several possible reasons. In principle, the power of the test should increase with increased sample size, increased allelic diversity, and more uniform allele frequencies.[60,61] The sample size (100 individuals or 200 haplotypes) is here limited for such highly polymorphic loci, and in the case of 'non-significant' association (as for example, for the pair A–Cw), we cannot positively state that there is no disequilibrium. In Table 2 it is also noted that, in general, the less polymorphic loci, here C4B, C4A, Bf and C2 (lowest number of alleles, lowest heterozygosity), present the lowest number of significant linkage disequilibrium with other loci (2, 2, 2, and 1, respectively). Although this relationship is not true for DPB1 and A (which are highly polymorphic but exhibit almost no significant linkage disequilibrium with other HLA loci), a significant correlation coefficient is verified in our data between the heterozygosity observed at a given locus and the number of significant linkage disequilibrium that this locus shares with other loci ($r = 0.60$, $P < 0.05$ given by the Student test when all loci are considered, and $r = 0.71$, $P < 0.01$ when only locus A, which is physically very distant, is removed). The scant information from some polymorphisms may thus partly explain the non-significant linkage disequilibrium found between some physically very close loci of the class III region.

Conclusions

The present study has allowed us to propose a linkage disequilibrium map of 14 loci of the MHC region. This map supports the hypothesis that certain parts of the MHC are prone to recombination hotspots, as in the case of the centromeric DP–DQ interval where we did not detect any linkage disequilibrium. On the other hand, other segments would be rarely fragmented (coldspots or frozen blocks), like the MHC telomeric part (locus A to B) which is characterised by significant linkage disequilibrium and a low recombination rate. It is noteworthy that the linkage disequilibrium observed in this segment would even extend beyond HLA–A towards the telomere, at least for the particular A1–B8 haplotype,[62] and thus include the postulated site of mutation for haemochromatosis.[63,64]

Both statistically significant linkage disequilibrium among distant loci and the prevalence of some extended haplotypes may be the result of directional selection, such as resistance to some selective pressure. However, some form of balancing selection also operates on HLA, since a lower homozygosity than expected is shown for several loci. Heterozygous advantage at MHC has mostly been explained by a more efficient response to various pathogens, but other mechanisms have also been invoked. Those involve frequency-dependent selection (advantage of rare alleles to which infectious agents are still unadapted), or fluctuating selection (turnover of advantageous alleles in response to epidemics).[65] These different kinds of selection may have acted simultaneously, or at different periods, to explain the complex pattern of HLA diversity in which both linkage disequilibrium and heterozygous advantage are currently observed.

Of course, the linkage disequilibrium pattern observed in this study only relates to the French population considered here. Different linkage disequilibrium maps may be found in other populations, because the creation, maintenance, or decay of linkage disequilibrium depend on several factors related to population dynamics, such as migrations and admixture between several populations having distinct gene frequencies, consanguinity levels, and genetic drift,[66] but also demographic stasis or expansion.[28] The conclusions drawn from the test for selective neutrality may also vary from population to population, as different selective pressures may influence the evolution of different populations. An accurate interpretation of both linkage disequilibrium and possible selective effects thus presupposes that extensive information is gathered at the population level for each studied data set, as was previously done, for example, in the *Provinces Françaises*[67] and the 12th HLA[5] workshops. Although a larger sample size than 100 individuals would be desirable for a more precise description of the French HLA diversity, the French origin of the families analysed in this

study has been checked over several generations, and the population is shown to be in Hardy-Weinberg equilibrium. This data set may thus be considered a representative sample for further analysing the effect of the history and dynamics of this population on both gene frequency distributions and linkage disequilibrium patterns across Europe.

Acknowledgements

This work was supported by the Swiss FNRS grants 31-49771.96 to AS-M and 32-047053.96 to LE and by a grant from the Société de Secours des Amis des Sciences (Paris VI) to SD and La Ligue Nationale Contre le Cancer. We thank Marie Legrand and Lucien Cazes for their expert technical assistance, and anonymous reviewers for their useful comments.

References

1 Campbell RD, Trowsdale J: The human major histocompatibility complex. *Immunol Today* 1997; **18**: 1.
2 Bodmer JG, Marsh SGE, Albert ED *et al*: Nomenclature for factors of the HLA system, 1996. *Tissue Antigens* 1997; **49**: 297–321.
3 Dausset J, Legrand L, Lepage V *et al*: A haplotype study of HLA complex with special reference to the HLA-DR series and to Bf, C2, and Glyoxylase I polymorphisms. *Tissue Antigens* 1978; **12**: 297–307.
4 Degos L, Dausset J: Human migrations and linkage disequilibrium of HLA system. *Immunogenet* 1974; **3**: 195–210.
5 Bodmer J, Cambon-Thomsen A, Hors J, Piazza A, Sanchez-Mazas A: Anthropology report: introduction. In: Charron D (ed). *Genetic Diversity of HLA: Functional and Medical Implication*. 12th International Histocompatibility Workshop and Conference, Paris, June 1996. EDK: Paris, 1997, vol 1, pp 269–284.
6 Imanishi T, Akaza T, Kimura A, Tokunaga K, Gojobori T: Estimation of allele and haplotype frequencies for HLA and complement loci. In: Tsuji K, Aizawa M, Sasazuki T (eds). *HLA 1991*. Oxford University Press: Oxford, 1992, vol 1, pp 76–79.
7 Abraham LJ, Leelayuwat C, Grimsley G *et al*: Sequence differences between HLA-B and TNF distinguish different MHC ancestral haplotypes. *Tissue Antigens* 1992; **39**: 117–121.
8 Degli-Esposti MA, Abraham LJ, McCann V, Spies T, Christiansen FT, Dawkins RL: Ancestral haplotypes reveal the role of the central MHC in the immunogenetics of IDDM. *Immunogenet* 1992; **36**: 345–356.
9 Hedrick PW: Gametic disequilibrium measures: proceed with caution. *Genetics* 1987; **117**: 331–341.
10 Lewontin RC: On measures of gametic disequilibrium. *Genetics* 1998; **120**: 849–852.
11 Begovich AB, McClure GR, Suraj VC *et al*: Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J Immunol* 1992; **148**: 249–258.
12 Carrington M, Stephens JC, Klitz W, Begovich AB, Erlich HA, Mann D: Major histocompatibility complex class II haplotypes and linkage disequilibrium values observed in the CEPH families. *Hum Immunol* 1994; **41**: 234–240.
13 Klitz W, Stephens C, Grote M, Carrington M: Discordant patterns of linkage disequilibrium of the peptide-transporter loci within the HLA class II region. *Am J Hum Genet* 1995; **57**: 1436–1444.
14 Cullen M, Erlich H, Klitz W, Carrington M: Molecular mapping of a recombination hotspot located in the second intron of the human TAP2 locus. *Am J Hum Genet* 1995; **56**: 1350–1358.
15 Cullen M, Noble J, Erlich H *et al*: Characterization of recombination in the HLA class II region. *Am J Hum Genet* 1997; **60**: 397–407.
16 Martin M, Mann D, Carrington M: Recombination rates across the HLA complex: use of microsatellites as a rapid screen for recombinant chromosomes. *Hum Mol Genet* 1995; **4**: 423–428.
17 Moonsamy PV, Klitz W, Tilanus MG, Begovich AB: Genetic variability and linkage disequilibrium within the DP region in the CEPH families. *Hum Immunol* 1997; **58**: 112–121.
18 Mittal KK, Mickey MR, Singal DP, Terasaki PI: Serotyping for homotransplantation. Refinement of microdroplet lymphocyte cytotoxicity test. *Transplant* 1968; **6**: 913–927.
19 Colombani J, Hors J: Technique de recherche des anticorps lymophocytotoxiques et determination des groupes tissulaires HLA. In: Allagille D (ed). *Techniques en Hématologie*. Flammarion Médecine-Sciences: Paris, 1972, pp 251–254.
20 Voorter CEM, Rozemuller EH, de Bruyn-Geraets D, van der Zwan A-W, Tilanus MGJ, van der Berg-Loonen EM: Comparison of DRB sequence-based typing using different strategies. *Tissue Antigens* 1997; **49**: 471–476.
21 Ueda A, Kusabat T, Yanase T: Determination of the second component of C' by electroimmunoassay *in sera* from patients with systemic lupus erythematosis. *Clin Exp Immunol* 1983; **53**: 217–224.
22 Gaither TA, Alling DW, Franck MM: A new one-step method for the functional assay of the fourth component (C4) of human and guinea pig complement. *J Immunol* 1974; **113**: 574–583.
23 Alper CA, Boenisch T, Watson L: Genetic polymorphism in human glycine rich B glycoprotein. *J Exp Med* 1972; **135**: 68–80.
24 Nedospasov SA, Udalova IA, Kuprash DV, Turetskaya RL: DNA sequence polymorphism at the human tumor necrosis factor (TNF) locus. Numerous TNF/lymphotoxin alleles tagged by two closely linked microsatellites in the upstream region of the lymphotoxin (TNF-beta) gene. *J Immunol* 1991; **147**: 1053–1059.
25 Jongeneel CV, Briant L, Udalova IA, Sevin A, Nedospasov SA, Cambon-Thomsen A: Extensive genetic polymorphism in the human tumor necrosis factor region and relation to extended HLA haplotypes. *Proc Natl Acad Sci USA* 1991; **88**: 9717–9721.
26 Guo S, Thompson E: Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 1992; **48**: 361–372.
27 Raymond M, Rousset F: An exact test for population differentiation. *Evolution* 1995; **49**: 1280–1283.
28 Slatkin M: Linkage disequilibrium in growing and stable populations. *Genetics* 1994; **137**: 331–336.
29 Weir BS: *Genetic Data Analysis II. Methods for Discrete Population Genetic Data*. 2nd edn. Sinauer Associates: Sunderland, Massachusetts, 1996.
30 Watterson GA: The homozygosity test of neutrality. *Genetics* 1978; **88**: 405–417.
31 Watterson GA: The homozygosity test after a change in population size. *Genetics* 1986; **112**: 899–907.
32 Ewens WJ: The sampling theory of selectivity neutral alleles. *Theoret Pop Biol* 1972; **3**: 87–112.
33 Schneider S, Kueffer J-M, Roessli D, Excoffier L: *Arlequin: a Software Package for Population Genetics*, vers. 1.0. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva: Geneva, 1997.
34 Terasaki PI, Gjertson DW (eds). *HLA 1997*.UCLA Tissue Typing Laboratory: Los Angeles, 1997.
35 Bunce M, Barnardo MCNM, Procter J, Marsh SGE, Vilches C, Welsh KI: High resolution of HLA-C typing by PCR-SSP: identification of allelic frequencies and linkage disequilibrium in 604 unrelated random UK Caucasoids and a comparison with serology. *Tissue Antigens* 1997; **50**: 100–111.
36 Crouau-Roy B, Briant L, Bouissou C *et al*: Tumor necrosis factor microsatellites in four European populations. *Hum Immunol* 1993; **38**: 213–216.
37 Gallagher G, Eskdale J, Oh H-H, Richards SD, Campbell DA, Field M: Polymorphisms in the TNF gene cluster and MHC serotypes in the West of Scotland. *Immunogenet* 1997; **45**: 188–194.
38 Sullivan KA, Wolfe MA, Lopez M, Jaspan JB, Bryer-Ash M: First report of recombination between the HLA-DR and HLA-DQ loci within a family. *Hum Immunol* 1997; **57**: 37–43.
39 Thomsen M, Alcalay D, Barmada M *et al*: MHC recombinant families. In: Charron D (ed). *Genetic Diversity of HLA: Functional and Medical Implications*. EDK: Sèvres, 1997, vol I, pp 188–195.

40 Robbins F, Hurley CK, Tang T *et al*: Diversity associated with the second expressed HLA-DRB locus in the human population. *Immunogenet* 1997; **46**: 104–110.
41 Trachtenberg EA, Erlich HA, Rickards O, DeStefano GF, Klitz W: HLA class II linkage disequilibrium and haplotype evolution in the Cayapa Indians of Ecuador. *Am J Hum Genet* 1995; **57**: 415–424.
42 Howell WM, Evans PR, Devereux SA, Sage DA, Smoth JL, Haegert DG: Absence of strong HLA-DR/DQ-DP linkage disequilibrium in the British and French Canadian Caucasoid populations. *Europ J Immunogenet* 1993; **20**: 363–371.
43 Jorde LB, Watkins WS, Carlson M *et al*: Linkage disequilibrium predicts physical distance in the adenomatous polyposis coli region. *Am J Hum Genet* 1994; **54**: 884–898.
44 Peterson AC, Di Rienzo A, Lehesjoki A-E, de la Chapelle A, Slatkin M, Freimer B: The distribution of linkage disequilibrium over anonymous genome regions. *Hum Mol Genet* 1995; **4**: 887–894.
45 Christiansen FT, Tay G, Smith LK *et al*: Histocompatibility matching for bone marrow transplantation donor-recipient pairs in the 4AOHW cell panel. *Hum Immunol* 1993; **38**: 42–51.
46 Dawkins RL, Zhang WJ, Degli-Esposti MA, Abraham LJ, McCann V, Christiansen PT: Studies of haplotypes by pulsed gel electrophoresis. In: Tait B, Harrison L (eds). *Baillière's Clinical Endocrinology and Metabolism, The Genetics of Diabetes*. Baillière Tindall: London, 1991, pp 285–298.
47 Strachan T, Read AP: *Human Molecular Genetics*. J Wiley: New York, 1996.
48 Nepom GT, Erlich H: MHC class-II molecules and autoimmunity. *Annu Rev Immunol* 1991; **9**: 493–525.
49 Izaabel H, Garchon H-J, Beaurin G *et al*: Distribution of HLA class II alleles and haplotypes in insulin-dependent Moroccan diabetics. *Hum Immunol* 1996; **49**: 137–143.
50 Hughes AL, Nei M: Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988; **335**: 167–170.
51 Hughes AL, Nei M: Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 1989; **86**: 958–962.
52 Hughes AL, Hughes MK: Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenet* 1995; **42**: 233–243.
53 Satta Y, O'hUigin C, Takahata N, Klein J: Intensity of natural selection at the major histocompatibility complex loci in primates. *Proc Natl Acad Sci USA* 1994; **91**: 7184–7188.
54 Takahata N, Nei M: Allelic geneaology under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 1990; **124**: 967–978.
55 Takahata N: Trans-species polymorphism of HLA molecules, founder principle, and human evolution. In: Klein J, Klein D (eds). *Molecular Evolution of the Major Histocompatibility Complex*. Springer-Verlag: Heidelberg, 1991, pp 29–49.
56 Klitz W, Thomson G, Baur MP: Contrasting evolutionary histories among tightly linked HLA loci. *Am J Hum Genet* 1986; **39**: 340–349.
57 Tiercy J-M, Sanchez-Mazas A, Excoffier L *et al*: HLA-DR polymorphism in a Senegalese Mandenka population: DNA oligotyping and population genetics of DRB1 specificities. *Am J Hum Genet* 1992; **51**: 592–602.
58 Sanchez-Mazas A, Dard P, Excoffier L, Langaney A: Selective neutrality tested on HLA. Abstract. *Eur J Immunogenet* 1995; **22**: 131.
59 Grundschober C, Sanchez-Mazas A, Excoffier L, Langaney A, Jeannet M, Tiercy J-M: HLA-DPB1 DNA polymorphism in the Swiss population: linkage disequilibrium with other HLA loci and population genetic affinities. *Eur J Immunogenet* 1994; **21**: 143–157.
60 Olson JM, Wijsman EM: Design and sample-size considerations in the detection of linkage disequilibrium with a disease locus. *Am J Hum Genet* 1994; **55**: 574–580.
61 Ott J, Rabinowitz D: The effect of marker heterozygosity on the power to detect linkage disequilibrium. *Genetics* 1997; **147**: 927–930.
62 Worwood M, Raha Chowdhury R, Robson KJH, Pointon J, Shearman JD, Darke C: The HLA A1-B8 haplotype extends 6 Mb beyond HLA-A: associations between HLA-A, B, F and 15 microsatellite markers. *Tissue Antigens* 1997; **50**: 521–526.
63 Feder JN, Gnirke A, Thomas W *et al*: A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996; **13**: 399–409.
64 Merryweather-Clarke AT, Pointon JJ, Shearman JD, Robson KJ: Global prevalence of putative haemochromatosis mutations. *J Med Genet* 1997; **34**: 275–278.
65 Hill AVS, Sanchez-Mazas A, Barbujani G *et al*: Human genetic variation and its impact on public health and medicine. In: Stearns SC (ed). *Evolution in Health and Disease*. Oxford University Press: Oxford, 1999, pp 62–74.
66 Hartl DL, Clark AG: *Principles of Population Genetics*, 2nd edn. Sinauer Associates: Sunderland, Massachusetts, 1989.
67 Ohayon E, Cambon-Thomsen A (eds): *Génétique des Populations Humaines* (Human Population Genetics). INSERM: Paris, 1986.