ELSEVIER

High Resolution HLA-DRB1 Identification of a Caucasian Population

Fionnuala Williams, Ashley Meenagh, Rich Single, Mark McNally, Philip Kelly, Mark P. Nelson, Diogo Meyer, Alex Lancaster, Glenys Thomson, and Derek Middleton

ABSTRACT: Polymerase chain reaction–sequence-specific oligonucleotide probes typing methods have been applied to 1000 individuals from the Northern Ireland population to give human leuckocyte antigen DRB1 (HLA-DRB1) allele assignment. HLA-DRB1 allele frequencies and four-locus haplotypes (A/B/C/DR) for this Caucasian population, based on HLA class I and class II allele assignment, are now presented. No significant deviations from Hardy-Weinberg proportions were observed. The HLA-C locus exhibited marginal evidence of selection ($p < 0.03$, uncorrected one-sided test) in the direction of balancing selection; the HLA-A, -B, and -DRB1 allele frequency distributions were compatible with expectations under a neutral model (which does not mean that selection is not operating). Evidence for selection was seen on haplotypes HLA-A*010101-B*0801-DRB1*030101 and HLA-A*290201-B*440301-DRB1*070101 based on their patterns of linkage disequilibrium. *Human Immunology 65, 66–77 (2004).*

KEYWORDS: HLA-DRB1; high resolution; sequence-specific oligonucleotide probes

ABBREVIATIONS

PCR	polymerase chain reaction
SSOP	sequence-specific oligonucleotide probes
HWP	Hardy-Weinberg proportions
EM	expectation maximization
LD	linkage disequilibrium
DPA	disequilibrium pattern analysis
CDV	constrained disequilibrium values

INTRODUCTION

Molecular identification of the alleles of human leukocyte antigen (HLA) class I and II genes is now standard practise in many HLA laboratories, resulting in improved accuracy and greater cost efficiency when compared with conventional serologic methods. Previously we have used polymerase chain reaction (PCR) and sequence-specific oligonucleotide probe (SSOP) methods to determine the allele frequencies at HLA-A, -B, and -C loci in 1000 Caucasian individuals from Northern Ireland [1–3]. We have now implemented a two-stage or three-stage PCR-SSOP typing approach for HLA-DRB1 allele identification. Amplification and probing within the first-stage results in allelic family definition, with second (or further) rounds of amplification and probing performed, for allele definition, based upon the first stage assignment. This enabled estimation of the four locus haplotypes (at the allele level) present in this Western European population.

MATERIALS AND METHODS

Population Sample

Originally Ireland consisted of four provinces, Connaught, Leinster, Munster, and Ulster, with Ulster divided into nine counties. Today, Northern Ireland is

From the Northern Ireland Regional Histocompatibility and Immunogenetics Laboratory (F.W., A.M., M.M., P.K., D.M.), City Hospital, Belfast, Northern Ireland; Department of Medical Biostatistics (R.S., M.P.N.), University of Vermont, Burlington, VT, USA; Department of Integrative Biology (D.M., A.L., G.T.), University of California, Berkeley, CA, USA; Queens University (D.M.), Belfast, Northern Ireland; and University of Ulster (D.M.), Coleraine, Northern Ireland.

Address reprint requests to: Dr. Fionnuala Williams, Northern Ireland Regional Histocompatibility and Immunogenetics Laboratory, Blood Transfusion Building, City Hospital, Belfast BT9 7TS, Northern Ireland; E-mail: fionnuala.williams@bll.n-i.nhs.uk.

Received July 8, 2003; revised September 29, 2003; accepted October 7, 2003.

0198-8859/04/$–see front matter
doi:10.1016/j.humimm.2003.10.004

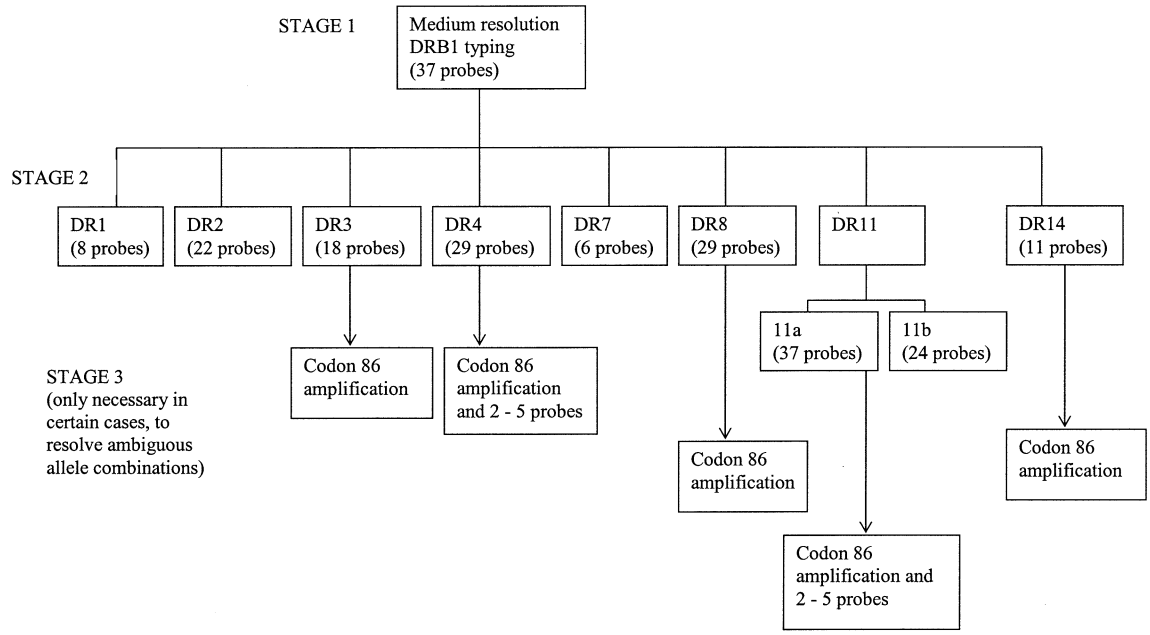

FIGURE 1 Polymerase chain reaction–sequence-specific oligonucleotide probe typing strategy for allele discrimination. Allele discrimination using codon 86 amplifications (stage 3) are only necessary in certain situations. Allele resolution is achieved by amplification alone in most cases, but occasionally it may be necessary to challenge the resulting amplicons with a small number of probes to obtain the desired allele result.

comprised of six of the original Ulster counties. Due to the close proximity of the north of Ireland and Scotland, migrations between these countries have been constant. The great settlement of Scottish and Northern English populations took place in the 17th century during the plantation of Ulster by the British Crown. Later, many of these Scottish "planter" families moved to America, under the title of "Scotch Irish." During the past 150 years there has been a steady, though small, infiltration from the United Kingdom mainland, arising mainly from commercial activity. Thus the population of Northern Ireland can be considered as comprising three traditions: the "original Irish," the "Ulster Scots," and the English.

The developed techniques were applied to 1000 Caucasian blood donors from Northern Ireland. All individuals had previously undergone HLA class I allele identification in similar PCR-SSOP high resolution typing studies [1–3].

PCR Amplification of the HLA-DRB1 Gene

Figure 1 illustrates the PCR-SSOP typing scheme for allele discrimination. Initially a set of seven primer sequences, which collectively would cover all HLA-DRB1 allele sequences (with the exception of HLA-DRB1*1130), have been selected from the beginning of exon 2 of the HLA-DRB1 gene. Combining these primers with an HLA-DRB generic reverse primer sequence, from the end of exon 2, yields an HLA-DRB1 specific product under selective amplification conditions. The increased specificity of this product permits the use of oligonucleotide probe sequences, which would produce conflicting hybridization signals for HLA-DRB1, -DRB3, -DRB4, and -DRB5 genes when challenged with the traditional HLA-DRB generically amplified PCR product [4]. Use of the developed medium resolution HLA-DRB1 PCR-SSOP typing strategy results in HLA-DRB1 allelic family resolution in the majority of cases.

Genomic DNA, previously isolated from whole blood using the salting out procedure [5], was subjected to PCR amplification in 150-μl reaction volumes. Each reaction contained: 67-mM Tris-HCl, pH 8.8; 16-mM $(NH_4)_2SO_4$; 1.5-mM $MgCl_2$; 0.01% (w/v) Tween; 200-μM of each dNTP; 2 units Taq polymerase (Bioline, London, United Kingdom); and 0.1- to 0.5-μg DNA. PCR primers (Table 1) were used at 0.2 μM (DR2, DR3, DR7, DR10, AMPBn) and 0.4 μM (AMP1, AMP4, DR9) within each reaction to yield locus specific products under specified cycling conditions (Table 1). Each PCR involved an initial denaturation of 96 °C for 5 minutes and the final PCR cycle was extended by an additional 5-minute incubation at 72 °C. Amplification was verified by 1.5% agarose gel electrophoresis. HLA-DRB1 specificity was confirmed by challenging resulting amplicons with probe sequences specific to HLA-DRB3, -DRB4, and -DRB5 allele sequences,

TABLE 1 HLA class II primers used for HLA-DRB1 medium and high resolution PCR-SSOP typing

Amplified product	Primer name	5′ Sequence 3′	Codon position	Band size (bp)	HLA-DRB1 alleles amplified
HLA-DRB1 generic[a]	AMP 1	TTC TTG TGG CAG CTT AAG TT	exon 2 7–13	270 (approx.)	All HLA-DRB1 alleles (except DRB1*1130)
	DR 2	TTC CTG TGG CAG CCT AAG AGG	exon 2 7–13		
	DR 3	CCA CGT TTC TTG GAG TAC TC	exon 2 5–11		
	AMP 4	GT TTC TTG GAG CAG GTT AAA C	exon 2 6–12		
	DR 7	A CGT TTC CTG TGG CAG GG	exon 2 5–10		
	DR 9	CGT TTC TTG AAG CAG GAT AAG TT	exon 2 6–13		
	DR 10	ACC AGA CCA CGT TTC TTG GAG G	exon 2 3–10		
	AMPBn	CCG CTG CAC (T/C)GT GAA (G/T)CT CT	exon 2 93–87		
HLA-DR1[a]	AMP 1	as above	as above	261	HLA-DRB1*01
	AMPBn	as above	as above		
HLA-DR2[a]	DR 2	as above	as above	261	HLA-DRB1*15 and *16
	AMPBn	as above	as above		
HLA-DR3 group[b]	3/11/6 GF	GT TTC TTG GAG TAC TCT ACG TC	exon 2 6–13	232	HLA-DRB1*03, *1107, *1127, *1333
	DR3R	GTA GTT GTG TCT GCA GTA (G/A)T	exon 2 83–77		
HLA-DR4[a]	AMP 4	as above	as above	263	HLA-DRB1*04, *1122, *1410
	AMPB	CCG CTG CAC TGT GAA GCT CT	exon 2 93–87		
HLA-DR7[a]	DR 7	as above	as above	265	HLA-DRB1*07
	AMPBn	as above	as above		
HLA-DR8 group[b]	8/12 GF	GT TTC TTG GAG TAC TCT ACG GG	exon 2 6–13	263	HLA-DRB1*08, *12, *1105, *1317, *14 (04/11/15/28/31)
	AMPBn	as above	as above		
HLA-DR11 group[c]	3/11/6 GF	as above	as above	232	HLA-DRB1*11, *13, *14 (alleles not amplified within the other systems)
	11/6R	GTA GTT GTG TCT GCA GTA GG	exon 2 83–77		
HLA-DR14 group[d]	3/11/6 GF	as above	as above	234	HLA-DRB1*14 (certain alleles), *1113, *1117
	DR14R	CC GTA GTT GTG TCT GCA A	exon 2 84–78		

The PCR cycling conditions for these primer combinations are as follows:

[a] 96 °C/1 min, 57 °C/1 min, 72 °C/1 min (38 cycles);

[b] 96 °C/1 min, 64 °C/1 min, 72 °C/1 min (15 cycles), then 96 °C/1 min, 56 °C/1 min, 72 °C/1 min (25 cycles);

[c] 96 °C/1 min, 65 °C/1 min, 72 °C/1 min (15 cycles), then 96 °C/1 min, 62 °C/30 sec, 72 °C/1 min (25 cycles);

[d] 96 °C/1 min, 72 °C/1 min (38 cycles).

Abbreviations: HLA = human leukocyte antigen; PCR = polymerase chain reaction; SSOP = sequence-specific oligonucleotide probe.

respectively, ensuring no coamplification of these other HLA-DRB loci.

High resolution allele discrimination initially involved eight different group specific amplifications (Figure 1), performed in 100-μl reaction volumes containing reagents described above. PCR primers, for each of the groups (Table 1) were used at a concentration of 0.2 μM in each reaction. These primer combinations were designed to specifically amplify only the HLA-DRB1 alleles within each of the groups (Table 1). PCR product specificity was verified by ensuring no amplification of alleles that were closest in sequence to each 5′ end primer.

Hybridization and Chemiluminescent Detection

Amplified DNA (2 μl) was applied to a charged nylon membrane as previously reported [6]. Hybridization with digoxigenin-labeled probes (Table 2) and chemiluminescent detection procedures were as previously described [1]. The probes were derived from polymorphic sequences within exons 2 of the HLA-DRB1 gene. Probe conditions (Table 2) were established by ensuring correct hybridization patterns with known control samples. Expected reaction patterns for probes used in the HLA-DRB1 typing systems are available on request from the authors. HLA-DRB1 allele resolution was based upon HLA-DRB1 alleles identified as of December 1999 [7].

Due to the complexity of the HLA-DRB1*13 allelic families, amplified within the HLA-DR11 subtyping group, the group was divided into two separate probing systems (11a and 11b) according to the polymorphism at codon position 47 (*i.e.*, TTC or TAC). HLA-DRB1*13 positive individuals were typed, in the first instance, with the HLA-DRB1*11a group of probes, which are used for the identification of the majority of HLA-DRB1*13 alleles including HLA-DRB1*130101 and -DRB1*130201, common HLA-DRB1*13 alleles in Caucasoid populations. Identification of HLA-

TABLE 2 Sequence-specific oligonucleotide probes used for HLA-DRB1 allele identification

Name	5′ Sequence 3′	Wash temperature concentration	Codon position
D21[e]	C CTG TGG CAG GGT AAA TAT	56 °C/20pm	7–13
1006[e]	TGG CAG GGT AAG TAT AAG	52 °C/20pm	9–14
1003[M]	G TAC TCT ACG TCT GAG TG	56 °C/10pm	9–15
1002[M]	AG CCT AAG AGG GAG TGT C	56 °C/20pm	10–16
DR18[M]	CT ACG GGT GAG TGT TAT	48 °C/20pm	11–16
D40[i]	GAG TGT CAA TTC TTC A	46 °C/20pm	14–19
D49[M]	TTC TTC AAC GGG ACG GAG	56 °C/30pm	17–22
D66[i]	G CGG GTG CAG TTC CTG	54 °C/20pm	22–27
D69[M,c,g]	G GTG CGG TAC CTG GAC A	58 °C/10pm	23–29
DR9[M]	GTG CGG TAT CTG CAC AGA G	58 °C/40pm	24–30
D72[b]	G CGG TTC CCG GAC AGA T	56 °C/20pm	24–30
DR25[M,c,g,h,i]	CGG TTC CTG GAC AGA TA	58 °C/75pm	25–30
2801[M]	CGG TTG CTG GAA AGA TGC	60 °C/20pm	25–30
D78c[b]	C CTG GAC AGA CAC TTC	50 °C/20pm	26–31
D78d[b,c]	C CTG GAC AGA TAC TTC	50 °C/20pm	26–31
D78e[e]	C CTG GAA AGT CTC TTC	48 °C/20pm	26–31
D78f[e]	C CTG GAA AGA CTC TTC	48 °C/20pm	26–31
D78[M,f]	A CTG GAG AGA CAC TTC C	52 °C/30pm	26–32
D78b[M,c,g,h]	C CTG GAG AGA TAC TTC C	52 °C/30pm	26–32
DR4[M]	TAC TTC TAT CAC CAA GA(G/A)G	52 °C/30pm	30–36
D93[c]	C CAT AAC CGG GAG GAG A	54 °C/20pm	31–37
D98[c,d,f,g,h]	AC CAA GAG GAG TAC GTG	56 °C/20pm	33–38
D99[M,f,g,h,i]	C CAG GAG GAG TTC GT	48 °C/30pm	33–38
D99a[d]	C CAA GAA GAG TAC GTG CGC	60 °C/20pm	33–39
D99b[g]	C CAA GAG GAG AAC GTG C	54 °C/20pm	33–39
D99c[g]	C CAA GAG GAG TTC GTG C	54 °C/20pm	33–39
DRB12[f]	CAG GAG GAG CTC CTG CGC	58 °C/2pm	34–39
DRB6[M,b,c,g,h,i]	CAG GAG GAG AAC GTG CGC	62 °C/5pm	34–39
D100[d,g]	CAA GAG GAG TCC GTG CGC	60 °C/20pm	34–39
D101[f,g]	AA GAG GAG GAC GTG CGC	60 °C/20pm	34–39
D102[b]	G GAG GAG TCC GTG CGC T	58 °C/20pm	34–40
D102a[c]	G GAG GAG TCC GTG CGC T	64 °C/20pm	34–40
D103b[g]	GAG GAG GAC TTG CG	46 °C/20pm	35–39
D103[f]	GAG GAG TTC CTG CGC T	54 °C/20pm	35–40
D106[g]	GAG AAC CTG CGC TTC G	52 °C/20pm	36–41
DR10[M]	G TAC GCG CGC TAC GAC A	56 °C/20pm	36–42
D108[d]	G TAC GTG CAC TTC GAC AGC G	64 °C/20pm	36–43
D124b[a]	AGC GAC GTG GGG GAG TAC	60 °C/20pm	42–47
D124a[a]	AGC GAC GTG AGG GAG TAC C	62 °C/20pm	42–48
D135[M,b,c,f,g,h]	G GAG TAC CGG GCG GT	56 °C/20pm	45–50
D135b[b,c,d,f,g,h,i]	G GAG TTC CGG GCG GTG A	62 °C/10pm	45–51
D136[g,h]	GAG TTC CTG GCG GTG	54 °C/20pm	46–50
D137[d]	AG TAC CGG GCG GTG ACG GA	64 °C/20pm	46–52
D139[b]	TTC CGG GCG GCG ACG G	58 °C/20pm	47–52
D139b[d]	TAC CGG GTG GTG ACG GAG	64 °C/20pm	47–52
D145[d]	GCG GTG ATG GAG CTG GG	64 °C/20pm	49–54
D162[M,c,g]	G CGG CCT GAT GAG GA	50 °C/20pm	54–59
D162b[i]	G CGG CCT GAT GCC GAG	56 °C/20pm	54–59
DR22[M,c,d,f,g,h]	CGG CCT AGC GCC GAG TA	58 °C/30pm	55–60
D163[f]	CGG CCT GTT GCC GAG TA	56 °C/20pm	55–60
D163b[g]	CGG CCT GTC GCC GAG T	56 °C/20pm	55–60
D164c[f]	GG CCT ATC GCC GAG TA	52 °C/20pm	55–60
D164b[f]	GG CCT GCT GCC GAG TA	52 °C/20pm	55–60
DRB14/1[M,c,f,h,i]	GG CCT GCT GCG GAG CAC T	64 °C/10pm	55–61
D164[d]	GG CCT GAT ACC GAG TAC TG	60 °C/20pm	55–61
D165[M,c,d,f,g,h,i]	G CCT GAT GCC GAG TAC TG	58 °C/30pm	55–61
D165b[d, h]	G CCT GAC GCT GAG TAC T	54 °C/20pm	55–61
5703[d,f,i]	G CCT GAT GAG GAG TAC TG	54 °C/10pm	55–61

Continued

TABLE 2 Sequence-specific oligonucleotide probes used for HLA-DRB1 allele identification *(Continued)*

Name	5′ Sequence 3′	Wash temperature concentration	Codon position exon 2 84–78
D168[d]	T GAT GCC CAG TAC TGG	50°C/20pm	56–61
D169[g]	GAT GAG GAC TAC TGG	46°C/20pm	57–61
D169a[g]	GAT GAG GAG TAC TGG	46°C/20pm	57–61
D171[f]	T GCG GAG CAC TGG AAC	52°C/20pm	57–62
D172[c,f]	GCC GAG TCC TGG AAC AGC	60°C/20pm	58–63
D173[g]	AG GAG CAC TGG AAC AGC	54°C/20pm	58–63
DRB all[M]	TGG AAC AGC CAG AAG GAC	56°C/20pm	61–66
D188[b]	GC CAG AAG AAC ATC CTG	52°C/20pm	63–68
D190[M,b,d,f,g,h]	CAG AAG GAC CTC CTG GA	54°C/50pm	64–69
D192[b,d,f,g,h]	G AAG GAC ATC CTG GAA G	54°C/15pm	64–70
D193[b]	AAG GAC TTC CTG GAG CAG	56°C/20pm	65–70
D194[M,b,d,f,g,h]	AG GAC TTC CTG GAA GAC	52°C/25pm	65–70
D194b[b]	AG GAC ATC CTG GAG CAG G	58°C/20pm	65–71
DR27[M,a,b,g,h]	TG GAG CAG GCG CGG	50°C/25pm	68–72
D203[d]	TG GAG CGG AGG CGG GC	58°C/20pm	68–73
D203b[g,h]	TG GAA GAC AGG CGG GC	54°C/20pm	68–73
7031[M,g,h]	CTG GAA GAC AAG CGG GCC G	60°C/15pm	68–74
DRB13[M,a,d,g,h]	TG GAA GAC GAG CGG GCC G	64°C/10pm	68–74
DR24[M,i]	AG CGG AGG CGG GCC GAG	64°C/20pm	69–74
D206[d]	AG CAG AGG CAG GCC GCG	60°C/20pm	69–74
DR28[M,b,g]	A GAC AGG CGC GCC G	52°C/30pm	69–74
7012[M,a,d,g,h]	AC CGC GGC CCG CCT CTG C	68°C/15pm	69–75
DR23[M,d,g,h]	AC CGC GGC CCG CTT CTG C	70°C/10pm	69–75
D207[b]	A GAC AGG GCC GCC GCG GT	64°C/20pm	69–75
D211[M,c,g]	AAG CGG GGC CGG GTG	54°C/20pm	71–75
D213[M,c]	G CGG GGC CAG GTG GA	56°C/30pm	71–76
DRB8[M,b,d,f,g,h]	G CGG GCC CTG GTG GAC AC	64°C/20pm	71–77
D214c[g]	CGG GCC GTG GTG GAC	60°C/10pm	72–76
D214[M,d,f,g]	CGG GCC GAG GTG GAC A	62°C/20pm	72–77
D216[d,g,h]	G GCC GCG GTG GAC ACC	64°C/20pm	72–77
D214b[b,f,h]	CGG GCC GCG GTG GAC A	62°C/10pm	72–78
7004[M,c,d]	GGC CGG GTG GAC AAC TAC	62°C/20pm	73–78
D222[e]	G GTG GAC AAT TAC TGC AG	54°C/20pm	74–80
D223[e]	GTG GAC ACC GTG TGC AGA	58°C/20pm	75–80
D225[f]	G GAC ACC TAT TGC AGA CA	56°C/20pm	76–81
D226[f]	GAC ACC TAC TGC AGA CA	52°C/20pm	77–81
D244[b]	AAC TAC GGA GTT GTG G	48°C/20pm	82–87
D247b[M,a,d,f]	TAC GGG GCT GTG GAG AG	56°C/25pm	83–88
D247[M,a,b,d,f]	TAC GGG GTT GTG GAG AG	56°C/20pm	83–88
D248[M,a,b,d,f]	AC GGG GTT GGT GAG AGC	56°C/30pm	83–88
D249[f]	C GGG GTT GTT GAG AGC	52°C/20pm	83–89

Of the 99 probe sequences employed for HLA-DRB1 typing, 41 probes were used in more than one typing system:

[M] SSOPs used in medium resolution HLA-DRB1 SSOP;

[a] SSOPs used for HLA-DR1 group;

[b] SSOPs used for HLA-DR2 group;

[c] SSOPs used for HLA-DR3 group;

[d] SSOPs used for HLA-DR4 group;

[e] SSOPs used for HLA-DR7 group;

[f] SSOPs used for HLA-DR8 group;

[g] SSOPs used for HLA-DR11a group;

[h] SSOPs used for HLA-DR11b group;

[i] SSOPs used for HLA-DR14 group.

Abbreviations: HLA = human leukocyte antigen; SSOP = sequence-specific oligonucleotide probes.

DRB1*130301, also present in Caucasians, is within the 11b group. Analysis with HLA-DRB1 medium resolution PCR-SSOP would always indicate if HLA-DRB1*130301 was likely to be present, dictating additional analysis within the 11b group for samples fitting into this category.

Codon 86 Resolution of Ambiguous HLA-DRB1 Allele Typing Results

Amplification, using primers (86AMP-GR and 86AMP-VR) designed around polymorphisms at codon 86 [4], are necessary for the separation of some alleles following PCR-SSOP high resolution typing (Figure 1). Depending on the allele combination to be resolved, these two reverse sequence primers would be combined with the forward primer, which is used for amplification of the allelic family (*e.g.*, AMP4 in the case of HLA-DRB1*04 alleles) in separate amplifications, using reagents outlined previously for HLA-DRB1 amplification. Cycling conditions were similar to those used for HLA-DR3 amplification (Table 1) except for annealing temperatures of 66 °C and 58 °C as appropriate.

Statistical Methods

HLA-DRB1 allele family (two digit) and allele (four/six digit) frequencies for the Northern Ireland population were determined by direct counting. To test for departures from Hardy-Weinberg proportions (HWP) an exact test was used [8], as implemented in the Arlequin software package [9], accessed via PyPop [10–12]. Standard chi-square analysis was also applied via PyPop to the set of common genotypes (expected values ≥ 5), as well as to individual genotypes, individual heterozygotes by allele (*e.g.*, DRB1*010101/X where X denotes all non-010101 alleles), all homozygotes, and all heterozygotes. To test the fit of the observed allele frequency distribution to that expected under neutrality (for the given sample size and observed number of alleles) the Ewens Watterson homozygosity test as implemented by Slatkin [13, 14] was applied via PyPop, using a one-tailed test that the observed homozygosity statistic was significantly less than expected under neutrality, *i.e.*, in the direction of balancing selection.

Combining the DRB1 typing results with previous HLA-A, -B, and -Cw allele assignments for this same group of 1000 individuals [1–3], HLA-A/B/C/DR four-locus haplotypes for the Northern Ireland population were estimated by the expectation-maximization (EM) algorithm using the EMHAPLOFREQ module in PyPop. Linkage disequilibrium (LD) was investigated at the individual coefficient level ($D_{ij} = f_{ij} - p_i q_j$, $D'_{ij} = D_{ij} / D_{max}$; where f_{ij} is the frequency of haplotype ij, and p_i and q_j are the frequencies of the constituent alleles) and using overall measures, D' and W_n, for each pair of loci [15, 16]. Note that D' without any subscript refers to the overall measure, whereas the subscripts indicate individual values. Significance of the overall LD was tested using the following permutation approach [12, 17]:

$$D' = \sum_{i=1}^{I} \sum_{j=1}^{J} p_i q_j |D'_{ij}|, \quad W_n = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} D_{ij}^2 / p_i q_j}{\min (I-1, J-1)}} \tag{1}$$

The disequilibrium pattern analysis (DPA) method was used to identify patterns of LD that are consistent with past selective events [18, 19]. Under selection in the recent past on the HLA class I two-locus haplotype HLA-A*010101-B*0801 (either directly on one or both of these alleles or via a hitchhiking event), then related haplotypes, *e.g.*, all the HLA-A*010101 non-B*0801 haplotypes, are predicted to have an expected value of LD proportional to the frequency of the unshared allele. We refer to these plots as A(010101): B plots, where A*010101 is the allele conditioned on. Figures illustrating fit to this pattern and lack of fit are given later. Fit to this prediction was based on a high correlation coefficient ($R^2 > 0.8$) for LD values in the negative space and few LD values in the positive space.

The constrained disequilbrium values (CDV) method [20–22] was used to detect selection events in the HLA region by examining the pattern of normalized LD values imposed by a three-locus system (D''_{ij}) compared to those in the respective two-locus system (D'_{ij}). The difference between these two measures, $\Delta = |D'_{ij} - D''_{ij}|$, has a distribution that can be indicative of selection (not all cases of selection will be detected by this method). The following criteria are used to infer selection based on Δ values:

1. If one of the three Δ values is positive (in practice greater than 0.1) and the remaining two are zero or negative, this is indication of selection;
2. If more than one of the Δ values is positive, but one is much larger than the rest (in practice more than double the next largest), this is indication of selection;
3. If all three Δ values are less than or equal to zero, or two are positive but close in value, no conclusion about selection can be drawn.

It was previously thought [21] that in cases 1 and 2, the constraining allele that gives the high positive Δ value is the one experiencing selection. Further studies [22] have shown that this can be misleading, especially when the center locus of three is the one leading to a high positive Δ value.

TABLE 3 HLA-DRB1 frequencies for the Northern Ireland population

HLA-DRB1*	Percent (n = 1000)	HLA-DRB1* alleles assigned	Percent (n = 1000)	Allele frequency
01	22.6[a]	010101	15.4	0.0790
		010201	1.6	0.0080
		0103	6.3	0.0330
03	29.9	030101	29.9	0.1600
04	33.9[a]	040101	20.9	0.1105
		0402	0.6	0.0030
		040301	0.7	0.0035
		0404	9.3	0.0470
		0405	0.7	0.0035
		0406	0.1	0.0005
		040701	3.4	0.0180
		0408	0.3	0.0015
		0416	0.1	0.0005
07	30.1	070101	30.1	0.1600
08	4.3	080101	3.4	0.0170
		080201	0.2	0.0010
		080302	0.4	0.0020
		080401	0.1	0.0005
		0806	0.2	0.0010
09	1.4	090102	1.4	0.0070
10	1.4	100101	1.4	0.0070
11	8.0[a]	110101	4.7	0.0235
		1102	0.2	0.0010
		1103	0.7	0.0035
		110401	2.5	0.0130
12	2.5	120101	2.5	0.0125
13	15.5[a]	130101	7.2	0.0365
		130201	6.4	0.0330
		130301	2.3	0.0115
14	2.9	140101	2.9	0.0145
15	33.6	150101	32.9	0.1820
		150201	0.7	0.0035
16	0.4	160101	0.4	0.0020

[a] The allele family percentage does not equal the total percentage of alleles identified for these allele families due to some individuals being heterozygous within the same family.

RESULTS

Amplification of HLA-DRB1, with the multiple HLA-DRB1 specific primers (Table 1), was optimal using PCR primers at concentrations of 0.2 μM or 0.4 μM, as appropriate. Initially all primers were employed at equal concentration (*i.e.*, 0.2 μM) but comparatively poor amplification of HLA-DRB1*01, -DRB1*04, and -DRB1*09 alleles was observed in the presence of other HLA-DRB1 alleles. As a consequence, various concentration titration experiments involving the primers amplifying these allele families (*i.e.*, AMP 1, AMP 4, and DR9) were conducted, with the combination of 0.4 μM for these three primers and 0.2 μM for all other primers proving optimal.

All 13 HLA-DRB1 allele families were identified, with allelic variation detected within 6 families (Table 3). No novel HLA-DRB1 allele sequences were encountered within the large group of samples studied. The HLA-DRB1 observed homozygosity frequency was 10.2% (expected under Hardy-Weinberg is 11.1%). Forty-three individuals homozygous at the allelic family level became heterozygous due to the presence of two alleles from the one family. This occurred in the allelic families of HLA-DRB1*01, -DRB1*04, -DRB1*11, and -DRB1*13. No evidence of deviation from HWP was seen either with the exact test or the overall chi-square test on common genotypes. At the individual genotype level HLA-DRB1*010101/070101 exhibited deviation from HWP (observed 37, expected 25.28, chi-square = 5.43, uncorrected $p < 0.02$) but not after correcting the type I error probability for even a modest number of the individual genotype tests (a total of 73 tests were performed). Application of the Ewens Watterson homozygosity test of neutrality indicated no evidence of deviation from neutrality with respect to the HLA-A, -B, or -DRB1 loci. The HLA-C locus revealed significant

TABLE 4 Estimates of HLA-A,- B,- C, and -DR allele haplotype copies occurring with frequencies ≥ 1% in the Northern Ireland population

Haplotype number	HLA-A*	HLA-Cw*	HLA-B*	HLA-DRB1*	Frequency
1	010101	070101	0801	030101	0.09007
2	03010101	07020101	070201	150101	0.04763
3	020101	0501	44020101	040101	0.03967
4	020101	07020101	070201	150101	0.03137
5	290201	1601	440301	070101	0.02056
6	020101	070101	0801	030101	0.01829
7	010101	0602	570101	070101	0.01786
8	020101	0501	44020101	150101	0.01061

deviation from neutrality ($p < 0.03$, one-sided test, uncorrected for multiple tests) in the direction of balancing selection.

Table 4 details the most frequently identified HLA-A/B/C/DR haplotypes in this Caucasian population. Note that the three-locus haplotype HLA-Cw*070101-B*0801-DRB1*030101 occurs on two common haplotypes, with -A*010101 (haplotype 1) and -A*020101 (haplotype 6); HLA-Cw*070201-B*070201-DRB1*150101 occurs with -A*03010101 (haplotype 2) and -A*020101 (haplotype 4); and HLA-A*020101-Cw*0501-B*44020101 occurs with -DRB1*040101 (haplotype 3) and -DRB1*150101 (haplotype 8). Of the alleles in these eight most common haplotypes, those with a frequency < 10% are HLA-A*290201 (4.85%), HLA-Cw*1601 (5.05%), HLA-B*440301 (6.45%) (all three occur on haplotype 5), and HLA-B*570101 (3.75%) (haplotype 7).

Overall LD was significant for each locus pair after correcting for the six pairwise tests. The strength of overall LD for each pair of loci is summarized in Table 5. The C:B pair demonstrates the highest LD by both measures, the B and C pairwise combinations with A and DRB1 exhibit lower values, and the A:DRB1 pair the lowest values, as expected based on their genetic distances. The D′ measure is known to have higher values than Wn [23]. This is especially true when there are large numbers of haplotypes estimated to exist as a single copy ("singletons"), a situation that results in higher D′ values. The percentage of haplotypes that are singletons ranged from a low of 23%, for A:C, to a high of 44%, for C:B. The total frequency of these singletons is low, with a range from 2% to 5%, indicating that although there may be many of them they do not comprise a large portion of the frequency spectrum.

For the DPA and CDV analyses, given the high LD between the B and C loci, and that both methods require a "reasonable" amount of genetic distance between the loci considered, we restricted our analyses to the A:B: DRB1 loci. Haplotypes 1, 2, 3, and 5 (Table 4) all reveal very high positive pairwise LD between all three pairs of loci. In the DPA analyses for haplotype 1 illustrated in Figure 2, strong patterns indicative of selection were seen for the HLA-A*010101-B*0801 and HLA-B*0801-DRB1*030101 haplotypes, with these haplotypes standing alone in the upper right quadrants of the A*010101:B and B*0801:DRB1 plots, respectively, and a nearly linear arrangement of the haplotypes in the negative LD space. Evidence of selection for these haplotypes was also seen when conditioned on the other allele for these pairs (*i.e.*, DPA plots of B*0801:A and DRB1*030101:B [results not shown]).

The HLA-A*010101-DRB1*030101 haplotype manifested a pattern consistent with selection when conditioned on the DRB1*030101 allele (lower left plot in Figure 2). However, the presence of a second high fre-

TABLE 5 Overall pairwise linkage disequilibrium

HLA loci	*Wn*	*D′*	*p* Value	Percent of singletons	Total frequency of singletons
A:C	0.305	0.472	< 0.0001	0.23	0.02
A:B	0.443	0.528	< 0.0001	0.33	0.04
A:DRB1	0.271	0.342	< 0.0001	0.26	0.03
C:B	0.765	0.948	< 0.0001	0.44	0.02
C:DRB1	0.358	0.577	< 0.0001	0.32	0.04
B:DRB1	0.379	0.645	< 0.0001	0.34	0.05

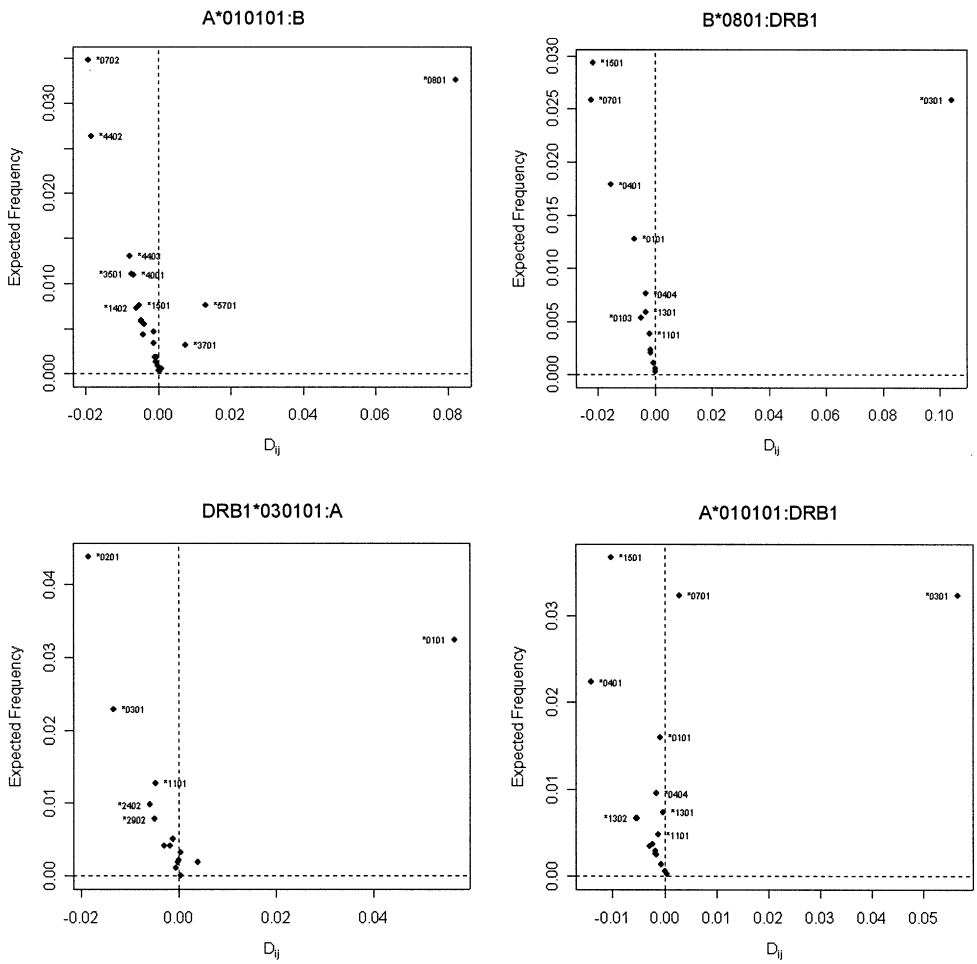

FIGURE 2 Disequilibrium pattern analysis for pairwise combinations of the A:B:DRB1 loci. All plots except A(010101):DRB1 reveal the pattern indicative of selection.

quency haplotype in the positive LD space and the scatter in the negative LD space (lower right plot in Figure 2) makes the evidence of selection less compelling when conditioned on the A*010101 allele.

For DPA conditioned on HLA-B*440301 (haplotype 5), a strong pattern of selection was seen with both the A and DRB1 plots with alleles HLA-A*290201 and -DRB1*070101 in the positive LD space (results not

TABLE 6 CDV analyses demonstrating evidence of selection

Constraint*	D	D′	D′′	Δ
Haplotype HLA-A*010101-B*0801-DRB1*030101				
A:B:(DRB1)	0.082	0.638	0.909	−0.271
A:(B):DRB1	0.057	0.444	0.091	+0.353
(A):B:DRB1	0.104	0.776	0.909	−0.173
Haplotype HLA-A*290201-B*440301-DRB1*070101				
A:B:(DRB1)	0.026	0.568	0.797	−0.229
A:(B):DRB1	0.014	0.358	0.201	+0.157
(A):B:DRB1	0.041	0.760	0.864	−0.104

* The constrained locus in each three-way combination is indicated in parentheses, *e.g.,* A:B:(DRB1) indicates that DRB1 is the constraining locus, so that D and D′ refer to the two-locus A:B haplotype, and D′′ refers to the two-locus A:B haplotype considered in the context of the three-locus A:B:DRB1 haplotype (with the alleles as indicated), and $\Delta = |D'_{ij} - D''_{ij}|$

shown). Both haplotypes 1 and 5 also exhibited strong evidence of selection with CDV analysis (Table 6). The following haplotypes did not reveal evidence of selection by CDV analysis (which does not mean selection is not acting) but inidicated strong patterns of selection by DPA analysis for the B*070201:A and DRB1*150101:B plots of haplotypes 2 and 4, and the B*570101:DRB1 plot of haplotype 7.

DISCUSSION

The HLA-DRB1 allele frequencies are based on the testing of 1,000 randomly selected individuals from the local bone marrow donor registry. Of the 33 HLA-DRB1 allelic variants identified (Table 3), 16 occur with frequencies $> 2\%$ in this Western European population. HLA-DRB1*04, the most frequently identified HLA-DRB1 allelic family in the population (33.9%), had nine allelic variants with HLA-DRB1*040101 predominating (20.9%). The HLA-DRB1*15 allelic family, close in frequency to HLA-DRB1*04, gave rise to the most frequently identified HLA-DRB1 allele in this population, HLA-DRB1*150101 (32.9%). Generally the results indicated that when more than one allele was identified within a given family, that one of these would be present at a much higher proportion than the others, *e.g.*, HLA-DRB1*110101 (4.7%), -DRB1*1102 (0.2%), -DRB1*1103 (0.7%), and -DRB1*110401 (2.5%). Only in the case of the HLA-DRB1*13 allele family was there a second allele present at a comparable frequency to the first allele. Three of the identified alleles were found in only one individual (HLA-DRB1*0406, -DRB1*0416 and -DRB1*08041). All of these alleles have previously been documented in Caucasian populations [24], with HLA-DRB1*0416 originally identified in this population [25].

The HLA-DRB1 allelic family resolution is greatly improved through the use of HLA-DRB1 specific PCR primers, compared with the traditional HLA-DRB amplification systems [4], minimizing the need for extended typing procedures, when allelic family identification only is required. Sole use of the first stage HLA-DRB1 SSOP typing system will result in certain ambiguous typing combinations, the majority of which involve a mixture of HLA alleles that would be expected in a Caucasian population, and other alleles that are not identified within our 1000 Caucasian individuals. This problem would also occur within low resolution HLA class I SSOP identification. Having access to HLA class I and HLA-DRB1 allele frequencies for this Caucasian population has enhanced the confidence of laboratory staff in assigning HLA types to individuals producing ambiguous typing information at a respective locus. When typing a local individual, allele combinations not identified in this population can be rendered unlikely.

The allele frequencies should give a good indication of the probability of finding a HLA-DRB1 allele matched donor for a patient requiring a bone marrow transplant, when dealing with a Caucasoid bone marrow donor registry. When bone marrow donors are selected on the basis of HLA-DRB1 matching at the allele family level, the allele frequency (Table 3) should give an indication of the likelihood of the recipient and donor being matched at the allele level, *e.g.*, if a HLA-DRB1*03 donor is selected as a HLA-DR match for a patient, then the chances of the pair being matched at the allele level are greater than if the donor and patient were matched for the HLA-DRB1*04 or - DRB1*13 allele family.

The haplotypes identified (Table 4) compare with those previously reported [26, 27], although they were at a serologic group level only and only for two- or three-locus haplotypes. Haplotypes 1 through 4 also achieved a frequency of at least 1% in data from the CEPH families [28]. Haplotypes 5 and 7 have a frequency just under 1% in the CEPH data. These six haplotypes are among the "multiple copy extended haplotypes" described by Bugawan and colleagues [28]. The corresponding frequencies for these six haplotypes from the CEPH data were lower than in the Northern Ireland population sample. However, direct comparison of the actual frequencies is complicated by the fact that haplotypes were inferred from family information in the CEPH data and estimated in the Northern Ireland data. Thus, we limit our remarks to general trends and the larger differences observed. Haplotypes 6 and 8 (Table 4) had substantially lower frequencies in the CEPH data. Additionally, three of the higher frequency haplotypes from the CEPH data had much lower frequencies in the in the Northern Ireland data and did not reach the 1% threshold (HLA-A*03010101-Cw*04010101-B*350101-DRB1*010101, HLA-A*020101-Cw*0602-B*570101-DRB1*070101, and HLA-A*020101-Cw*0602-B*1302-DRB1*070101). The Northern Irish frequencies and haploptyes provided conform to a Caucasian grouping. Although the data relates to a subsection of a European community, direct comparison with neighboring communities (*e.g.*, Southern Ireland, Scotland, England, and Wales) was not possible, because data of comparable resolution was not available.

Haplotype assignments and frequencies can prove beneficial in several clinical and research areas. Conserved extended HLA haplotypes may also indicate conservation within non-HLA loci, which becomes relevant in the case of matched unrelated bone marrow donor transplants, bringing haplotype matched transplants closer to the sibling transplant situation. Previous studies have discussed the possibility that haplotype frequencies might

exhibit ancestral relationships, whereas recent exchanges may be more apparent from allele frequencies [29]. Haplotype data highlights the diversity of human populations and has been found to be beneficial for distinguishing populations when allele frequency distributions are similar [30]. Additional to its use in anthropology studies, large sets of data of the type generated in this study are also useful as controls in disease studies. Haplotype associations would prove more informative especially where high degrees of LD renders it impossible to ascertain which single locus is involved in a disease.

Although the DPA and CDV analyses do not detect all cases of selection, they are useful tools to identify some cases of selection (probably more recent and strong selection events). In this study both the haplotypes HLA-A*010101-B*0801-DRB1*030101 (haplotype 1) and HLA-A*290201-B*440301-DRB1*070101 (haplotype 5) indicated evidence of selection by both methods. Both these haplotypes have been previously implicated as selected based on serologic data [19, 21]. It is not possible to determine which of the alleles on these haplotypes, or combinations of alleles, are the selected ones [22].

The main benefits of the data will only become apparent when similar sets of data are made available to researchers in the HLA field through channels such as International HLA workshops, journal publications, and websites. Recently a website "www.allelefrequencies.net" has been set-up to contain data such as presented in this study [31]. It is hoped that this will provide information in a convenient form to those working in the fields of anthropology, vaccine development, and stem cell transplantation.

ACKNOWLEDGMENTS

We thank the Northern Ireland Histocompatibility and Immunogenetics Fund for the financial support of F. Williams. The research of G. Thomson, M.P. Nelson, D. Meyer, A. Lancaster, and R. Single was kindly supported by NIH grants GM35326 and AI49213. We also thank the United Kingdom Transplant Support Service Authority (UKTSSA) for supplying cell-line DNA; staff at the Northern Ireland Histocompatibility and Immunogenetics laboratory for technical assistance; and Mrs A.M. McCann for assistance in typing the manuscript.

REFERENCES

1. Williams F, Meenagh A, Maxwell AP, Middleton D: Allele resolution of HLA-A using oligonucleotide probes in a two-stage typing strategy. Tissue Antigens 54:59, 1999.
2. Middleton D, Williams F, Hamill MA, Meenagh A: Frequency of HLA-B alleles in a Caucasoid population determined by a two-stage PCR-SSOP typing strategy. Hum Immunol 61:1285, 2000.
3. Williams F, Meenagh A, Patterson C, Middleton D: Molecular diversity of the HLA-C gene identified in a Caucasian population. Hum Immunol 63:602, 2002.
4. Bignon JD, Fernandez-Vina MA: Protocols of the 12th International Histocompatibility Workshop for typing of HLA class II alleles by DNA amplification by the polymerase chain reaction (PCR) and hybridization with sequence specific oligonucleotide probes (SSOP). In Charron D (ed): Genetic Diversity of HLA: Functional and Medical Implication. Paris: EDK, 1997.
5. Miller SA, Dykes DD, Poleskey HF: A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acid Res 16:1215, 1988.
6. Middleton D, Williams F, Cullen C, Mallon E: Modification of an HLA-B PCR-SSOP typing system leading to improved allele determination. Tissue Antigens 45:232, 1995.
7. Robinson J, Waller MJ, Parham P, Bodmer JG, Marsh SGE: IMGT/HLA database – a sequence database for the human major histocompatibility complex. Nucleic Acid Res 29:210, 2001.
8. Guo S, Thompson E: Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48: 361, 1992.
9. Schneider S, Roessli D, Excofier L: Arlequin: a software for population genetics data analysis. Version 2.000. Geneva, Switzerland: Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, 2000.
10. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G: PyPop: a software framework for population genomics: analysing large-scale multi-locus genotype data. In Altman RB, et al (eds): Pacific Symposium on Biocomputing 2003. River Edge, NJ: World Scientific, 2003.
11. Lancaster A, Nelson MP, Single RM, Meyer D, Thomson G: Software framework for the biostatistics core. In Hansen JA, Dupont B (eds.): HLA 2002. Immunobiology of the Human MHC. Seattle, WA: IHWG Press, 2003.
12. Single R, Meyer D, Thomson G: Statistical methods for analysis of population genetic data. In Hansen JA, Dupont B (eds.): HLA 2002. Immunobiology of the Human MHC. Seattle, WA: IHWG Press, 2003.
13. Slatkin M: An exact test for neutrality based on the Ewens sampling distribution. Genet Res 64:71, 1994.
14. Slatkin M: A correction to the exact test based on the Ewens sampling distribution. Genet Res 68:259, 1996.
15. Hedrick PW: Gametic disequilibrium measures: proceed with caution. Genetics 117:331, 1987.
16. Cramer H: Mathematical Methods of Statistics. Princeton, NJ: Princeton University Press, 1946.
17. Slatkin M, Excoffier L: Testing for linkage disequilibrium in genotypic data using the expectation-maximization algorithm. Heredity 76:377, 1996.
18. Thomson G, Klitz W: Disequilibrium pattern analysis. I. Theory. Genetics 116:623, 1987.

19. Klitz W, Thomson G: Disequilibrium pattern analysis. II. Application to Danish HLA-A and B locus data. Genetics 116:633, 1987.
20. Robinson WP, Asmussen M, Thomson G: Three locus systems impose additional constraints on pairwise disequilibria. Genetics 129:925, 1991.
21. Robinson WP, Cambon-Thomsen A, Borot N, Klitz W, Thomson G: Selection, hitchhiking and disequilibrium analysis at three linked loci with application to HLA data. Genetics 129:931, 1991.
22. Grote M, Klitz W, Thomson G: Constrained disequilibrium values and hitchhiking in a three-locus system. Genetics 150:1295, 1998.
23. Weiss KM, Clark AG: Linkage disequilibrium and the mapping of complex human traits. Trends Genet 18:19, 2002.
24. Marsh SGE, Parham P, Barber LD: The HLA Fact Book. New York: Academic Press, 2000.
25. Middleton D, Hughes DJ, Trainor F, Graham CA, Savage DA: An HLA-DRB1*04 first domain sequence (DRB1*0416) which differs from HLA-DRB1*0401 at codon 59. Tissue Antigens 43:44, 1994.
26. Schipper RF, Schreuder GMT, D'Amaro J, Oudshoorn M: HLA gene and haplotype frequencies in Dutch blood donors. Tissue Antigens 48:562, 1996.
27. Schipper RF, D'Amaro J, Bakker JT, Bakker J, van Rood JJ, Oudshoorn M: HLA gene and haplotype frequencies in bone marrow donors worldwide registries. Hum Immunol 52:54, 1997.
28. Bugawan TL, Klitz W, Blair A, Erlich HA: High resolution HLA class I typing in the CEPH families: analysis of linkage disequilibrium among HLA loci. Tissue Antigens 56:392, 2000.
29. Gibert M, Reviron D, Mercier P, Chiaroni J, Boetsch G: HLA-DRB1 and DQB1 polymorphisms in Southern France and genetic relationships with other Mediterranean populations. Hum Immunol 61:930, 2000.
30. Mack SJ, Bugawan TL, Moonsamy PV, Erlich JA, Trachtenberg EA, Paik YK, Begovich AB, Saha N, Beck HP, Stoneking M, Erlich HA: Evolution of Pacific/Asian populations inferred from HLA class II allele frequency distributions. Tissue Antigens 55:383, 2000.
31. Middleton D, Menchaca L, Rood H, Komerofsky R: New allele frequency database: www.allelefrequencies.net. Tissue Antigens 61:403, 2002.