3 Results

3.1 Explanatory Remarks and Definitions

The aim of this thesis is the development of a method, which allows the identification of HLA haplotypes as members of two groups, "frequent HLA haplotypes" and "rare HLA haplotypes". HLA haplotypes are the combinations of alleles of the different HLA genes on the same parental chromosome. The HLA genes extend over a region of around 4.5 Mbp. Frequent HLA haplotypes were selected by the five involved registries of the MADO project. Here the 15 most frequent HLA haplotypes identified in these registries were cumulated. In tables 23 to 27 these frequent HLA haplotypes, which were provided by the registries are presented.

FRANCE - FGM			
HLA-A	**HLA-B**	**HLA-DR**	**Frequency**
1	8	3	4,49%
29	44	7	2,90%
3	7	15	2,76%
2	44	4	2,23%
3	35	1	1,27%
2	7	15	1,22%
2	44	7	1,09%
2	62	4	0,99%
2	51	11	0,97%
1	57	7	0,89%
23	44	7	0,87%
2	44	13	0,82%
2	44	11	0,80%
2	18	11	0,76%
2	51	13	0,62%

Table 23: The 15 most frequent HLA haplotypes of the French donor registry

ENGLAND- Anthony Nolan			
HLA-A	**HLA-B**	**HLA-DR**	**Frequency**
1	8	3	5,62%
2	44	4	3,07%
29	44	7	1,82%
2	7	15	1,82%
1	57	7	1,39%
3	35	1	1,18%
2	8	17	0,86%
2	60	13	0,81%
24	7	15	0,80%
2	57	7	0,79%
11	35	1	0,78%
2	44	7	0,75%
2	60	4	0,68%
23	44	7	0,64%
30	13	7	0,64%

Table 24: The 15 most frequent HLA haplotypes of the British donor registry

THE NETHERLANDS			
HLA-A	HLA-B	HLA-DR	Frequency
1	8	3	8,07%
3	7	15	3,24%
2	7	15	2,50%
3	35	1	2,26%
2	62	4	2,06%
2	60	13	1,87%
2	44	4	1,21%
29	44	7	1,20%
2	8	17	0,99%
2	62	13	0,94%
24	7	15	0,83%
2	57	7	0,83%
1	57	7	0,80%
2	44	11	0,61%
23	44	7	0,57%

Table 25: The 15 most frequent HLA haplotypes of the Dutch donor registry

ITALY- IBMDR			
HLA-A	HLA-B	HLA-DR	Frequency
1	8	3	2,52%
2	18	11	2,15%
2	51	11	1,97%
24	35	11	1,68%
3	7	15	1,23%
30	13	7	1,14%
24	18	11	1,14%
3	35	1	1,14%
29	44	7	1,05%
30	18	17	0,98%
33	14	1	0,96%
2	44	11	0,86%
2	35	14	0,77%
2	51	13	0,74%
2	51	14	0,38%

Table 26: The 15 most frequent HLA haplotypes of the Italian donor registry

HUNGARY			
HLA-A	HLA-B	HLA-DR	Frequency
1	8	3	6,66%
2	18	11	3,45%
2	44	4	1,69%
3	7	15	1,45%
2	13	7	1,41%
23	44	7	1,26%
2	8	3	1,10%
2	51	11	1,10%
3	7	11	1,09%
3	35	1	1,03%
2	27	16	0,98%
2	7	15	0,89%
25	18	15	0,82%
2	44	16	0,80%
2	50	7	0,69%

Table 27: The 15 most frequent HLA haplotypes of the Hungarian donor registry

Since molecular haplotyping of several HLA genes is technically not possible yet (due to the distances between the genes that are too big) the aim was the identification of the alleles of individual HLA genes which the frequent HLA haplotypes are made up of. These alleles were defined as “frequent alleles” and all the other alleles as “rare alleles”. The frequent alleles are listed in table 28. HLA alleles are variations of sequences of one gene.

HLA-A	HLA-B	HLA-DRB1
HLA-A*0101	HLA-B*0702	HLA-DRB1*0101
HLA-A*0201	HLA-B*0801	HLA-DRB1*0301
HLA-A*0301	HLA-B*1302	HLA-DRB1*0401
HLA-A*2301	HLA-B*1501	HLA-DRB1*0701
HLA-A*2402	HLA-B*1801	HLA-DRB1*1101
HLA-A*2902	HLA-B*3501	HLA-DRB1*1104
HLA-A*3001	HLA-B*3503	HLA-DRB1*1302
HLA-A*3002	HLA-B*4001	HLA-DRB1*1501
	HLA-B*4402	
	HLA-B*4403	
	HLA-B*5101	
	HLA-B*5701	

Table 28: Frequent alleles of HLA-A, -B and –DRB1

The aim of the MADO project is to increase the efficiency of registration of potential haematopoietic stem cell donors. This goal should be reached by using a pre-screening strategy. A robust, reliable and easy-to-use molecular method is required. Given the cost and complexity of these HLA typing technologies a new method needed to be developed. The key component of most technologies in HLA typing aimed at the identification and dissection of polymorphisms in the MHC genes.

The method presented in this thesis, which was develop to perform a sufficient pre-screening is based on the GOOD assay. This technology is, as detailed described in section ("the GOOD assay"), a competitive assay with the potential for a high degree of automation. The GOOD assay was adapted for analysis of the polymorphisms of interest. These polymorphisms were selected after dissection of the HLA class I and class II genes sequences and examination for their informativity in terms of identification of frequent or rare HLA alleles.

3.2 Sequence Analysis and Marker Selection

The latest sets (updated Jan. 2005) of FASTA sequences of HLA-A, HLA-B and HLA-DRB1 were downloaded from a public HLA-Sequence Database (IMGT/HLA Sequence Database). This database is a specialist database for sequences of the human MHC. It includes the official sequences of the WHO HLA Nomenclature Committee for factors of the HLA-System and is part of the international **I**m**M**uno**G**ene**T**ics (IMGT) project. In addition to the sequences the database provides detailed information concerning the materials, from which the sequence was obtained, and data on the validation of the sequences. To date 1'972 allele sequences, including 349 HLA-A, 626 HLA-B, 182 HLA-C, 5 HLA-E, 2 HLA-F and 15 HLA-G class I alleles have been named. Three HLA-DRA, 470 HLA-DRB, 28 HLA-DQA1, 60 HLA-DQB1, 22 HLA-DPA1, 116 HLA-DPB1, 4 HLA-DMA, 6 HLA-DMB, 8 HLA-DOA and 8 HLA-DOB class II alleles were named.

This thesis focuses on the alleles of HLA-A, HLA-B and HLA-DRB1. These are the transplantation relevant HLA genes and therefore of most importance for donor registries.

For HLA typing generally exons 2 and 3 of the HLA class I genes and exon 2 of HLA class II genes are used. In exon 2 of HLA-DRB1 109 polymorphic bases (PBs) are known in a 270 bp fragment (from base 101 to base 370). This corresponds to a polymorphism rate of 2.48. For HLA-A 152 PBs and 170 PBs for HLA-B were extracted. Based on a length of 545 bp (from base 74 to base 619) for exon 2 and 3 it is a polymorphism rate of 3.59 for HLA-A and 3.21 for HLA-B. The polymorphic bases are listed in table 29 for HLA-DRB1, in table 30 for HLA-A, and in table 31 for HLA-B.

HLA-DRB1							
Positions	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**
109	Y	175	B	240	R	301	S
112	D	176	V	241	R	302	V
113	R	177	Y	246	R	303	B
115	B	178	D	250	Y	304	K
116	W	181	B	253	S	305	S
117	S	184	M	256	R	306	Y
118	N	186	Y	257	N	307	S
119	N	188	R	258	Y	308	N

122	N	189	R	259	R	309	S
123	R	192	R	260	V	316	R
124	N	195	R	261	B	317	M
125	N	196	N	262	S	318	Y
126	K	197	H	263	R	319	B
127	R	199	B	264	S	320	W
129	R	200	Y	265	Y	321	B
133	Y	203	R	266	M	337	S
135	W	204	S	278	R	339	R
144	Y	206	W	283	R	341	Y
146	S	220	V	286	H	344	D
150	S	225	S	289	Y	345	K
155	S	227	W	293	R	351	M
161	D	228	Y	294	R	357	R
164	W	229	Y	295	S	364	M
165	N	230	D	296	R	366	R
167	H	233	Y	297	S	370	R
169	S	235	K	298	R		
171	V	236	Y	299	V		
174	W	239	B	300	R		

Table 29: Polymorphic Bases of HLA-DRB1

HLA-A							
Positions	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**
78	Y	242	Y	362	K	488	M
81	V	243	K	363	R	489	R
90	S	256	S	367	Y	493	Y
92	R	257	D	368	N	494	V
97	W	259	V	372	Y	497	Y
98	H	261	V	376	S	498	Y
102	H	265	S	385	Y	502	V
104	S	268	M	391	K	503	R
105	S	270	H	392	R	506	R
106	R	271	R	395	R	517	D
108	S	275	R	396	M	519	V
113	S	278	S	397	Y	521	Y
121	M	282	S	399	M	523	M
123	Y	289	R	402	S	524	D
125	R	290	B	403	Y	526	B
126	R	292	V	404	R	527	N
127	R	294	Y	411	Y	530	M
142	K	297	R	412	S	532	K
144	M	299	H	413	R	538	Y
160	D	301	R	414	V	539	D
163	D	302	R	416	R	542	R
171	M	303	Y	418	N	545	Y
176	R	307	S	419	H	553	S
180	D	308	S	420	Y	555	K
194	S	311	Y	423	Y	557	S

200	R	313	S	426	M	559	N
203	D	314	Y	427	K	560	N
212	Y	317	D	445	K	561	S
214	Y	318	Y	448	Y	564	M
219	R	319	S	450	R	565	S
224	R	324	S	453	M	570	S
228	R	331	S	455	W	571	K
233	R	333	S	456	R	583	Y
235	R	341	M	463	M	589	R
238	R	345	K	468	Y	595	V
239	R	346	W	477	R	601	R
240	K	351	Y	480	R	616	R
241	S	355	V	485	D	618	B

Table 30: Polymorphic Bases of HLA-A

HLA-B							
Positions	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**	**Positions**	**Bases**
83	Y	243	D	322	B	483	S
89	S	244	K	337	K	485	M
91	Y	245	V	341	M	486	V
97	B	246	R	345	K	489	R
103	K	247	Y	353	Y	493	Y
105	Y	248	W	354	Y	499	W
106	R	255	S	355	H	500	Y
117	Y	256	S	356	K	502	Y
119	K	257	V	357	S	503	R
126	R	258	K	360	S	506	D
131	Y	259	R	361	D	512	K
134	S	261	S	362	N	524	K
141	Y	263	Y	363	B	526	V
142	N	266	R	365	Y	527	H
144	M	269	H	368	N	528	S
146	Y	270	S	369	H	538	B
159	Y	271	W	379	V	539	D
161	R	272	N	387	S	540	V
165	S	273	S	395	K	544	R
167	W	277	V	396	Y	545	Y
171	S	278	B	397	Y	546	Y
175	M	280	H	404	S	548	R
181	R	281	M	408	K	555	R
186	M	282	S	409	Y	557	R
193	R	283	R	411	Y	559	V
200	M	285	R	412	V	560	N
201	R	287	R	414	M	566	Y
204	R	289	R	416	R	570	B
205	R	292	K	418	N	571	K
206	N	293	R	419	H	572	S
207	K	295	H	420	H	577	M
209	V	299	W	425	R	582	R

210	K	301	D	430	R	583	Y
211	M	302	R	431	R	589	R
213	S	304	Y	435	R	594	Y
215	S	309	S	436	R	603	S
221	Y	311	H	445	R	605	M
222	R	312	S	453	B	610	S
226	R	313	S	461	Y	614	R
228	R	314	Y	463	M	616	K
234	R	317	K	474	M	618	K
238	R	319	S	477	S		
242	S	320	V	481	R		

Table 31: Polymorphic Bases of HLA-B

Based on these polymorphisms over 120000 alleles of HLA-DRB1, over 280000 alleles of HLA-A, and more than 370000 alleles of HLA-B would be theoretically possible.

Some of the HLA alleles have identical sequences over the relevant exons. These alleles are not separated by the common HLA typing methods such as SBT, RSCA, SSPO, etc. Thus these alleles are combined and coded to shorten the list of alleles. Alleles with identical sequences and their coded names are listed in the tables 32, 33 and 34. These tables were extracted and modified from "Exon Identities and Ambiguous Typing Combinations" (Release 2.80; Anthony Nolan Trust; January 2005).

Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Allele 6	Allele 7	Allele 8	Code in table
A*010101	A*0104N							A*010101/0104N
A*02010101	A*02010102L	A*020108	A*020111	A*0209	A*0243N	A*0266	A*0275	A*02G1
A*0207	A*0215N							A*0207/0215N
A*0211	A*0269							A*0211/0269
A*021701	A*021702							A*0217
A*03010101	A*03010102N	A*03010103						A*030101G1
A*110101	A*1121N							A*110101/1121N
A*2301	A*2307N							A*2301/2307N
A*24020101	A*24020102L	A*240203	A*2409N	A*2411N	A*2440N			A*24G1
A*240301	A*2433							A*240301/2433
A*29010101	A*29010102N							A*2901
A*300101	A*300102							A*300101/300102
A*300201	A*300202							A*300201/300202
A*680102	A*6811N							A*680102/6811N
A*7401	A*7402							A*7401/7402

Table 32: HLA-A alleles sequence identical over exon 2 and 3

Allele 1	Allele 2	Allele 3	Allele 4	Allele 5	Code in table
B*0705	B*0706				B*0705/0706
B*0801	B*0819N				B*0801/0819N
B*15010101	B*15010102N				B*150101G1
B*1512	B*1519				B*1512/1519
B*15170101	B*15170102				B*151701
B*180101	B*1817N				B*180101/1817N
B*270502	B*2713	B*270504			B*27G1
B*350101	B*3540N	B*3542			B*35G1
B*390101	B*390103				B*3901G1
B*400101	B*400102	B*4055			B*40G1
B*400201	B*4056				B*400201/4056
B*40060101	B*40060102				B*400601G1
B*44020101	B*4427	B*44020102S	B*4419N		B*44G1
B*4501	B*4507				B*4501/4507
B*47010101	B*47010102				B*47010101/47010102
B*4801	B*4809				B*4801/4809
B*510101	B*5132	B*5130	B*510105	B*5111N	B*51G1
B*8101	B*8102				B*8101/8102

Table 33: HLA-B alleles sequence identical over exon 2 and 3

Allele 1	Allele 2	Allele 3	Code in table
DRB1*030501	DRB1*030502		DRB1*030501/030502
DRB1*040501	DRB1*040503	DRB1*040504	DRB1*0405G1
DRB1*040701	DRB1*040703		DRB1*040701/040703
DRB1*080201	DRB1*080202		DRB1*080201/080202
DRB1*080401	DRB1*080404		DRB1*080401/080404
DRB1*080402	DRB1*080403		DRB1*080402/03
DRB1*110101	DRB1*110102		DRB1*110101/110102
DRB1*110401	DRB1*110402		DRB1*110401/110402
DRB1*110801	DRB1*110802		DRB1*110801/110802
DRB1*120101	DRB1*1206	DRB1*1210	DRB1*120101/1206/1210
DRB1*130301	DRB1*130302		DRB1*130301/130302
DRB1*140701	DRB1*140702		DRB1*140701/140702

Table 34: HLA-DRB1 alleles sequence identical over exon 2

3.2.1 Marker selection and description for screening of the alleles of the genes HLA-A and HLA-B

At the beginning of this project, a strategy of screening for frequent alleles based on the informativity of the individual polymorphic bases was chosen. Each polymorphic base was weighted individually. The aim was to find a set of

polymorphic bases, which allowed defining HLA alleles with a high degree of resolution. A selection of polymorphic bases was made that enables a distinction of frequent or rare alleles. These polymorphic bases were selected by their separation informativity. For example, in HLA-A position 98 is polymorphic and has three alleles, A, G and T. One third of all known HLA-A alleles have an A at this position, another third a G and the last third the T. None of the frequent alleles carry an A at this position. Thus the frequent alleles can be distinguished from 30 % of the rare alleles. It divides the group of frequent alleles into two groups. Some of the frequent alleles carry a G and some T at this position. With this quite good distinction between frequent and rare HLA-A alleles can be achieved. Position 98 is in this strategy of marker selection and for this specific question the most informative position of HLA-A. The next step was to identify the next most informative base. Step by step sets of polymorphic positions for HLA-A (16) and HLA-B (18), respectively, were selected. The problem with this strategy is that in HLA genes polymorphic bases are very close to each other. This makes it very difficult to create assays that take only one position into account.
On the other hand the proximity of the polymorphic bases has added potential. A combination of polymorphic bases and the phase of alleles of individual positions has a higher degree of informativity than a single base, since short sequence fragments, microhaplotypes are achieved. In some cases microhaplotypes are very specific, for example, the marker HLAB_272f1. For this marker 8 different microhaplotypes can be detected (ACAT, ACTG, CCTT, AGTA, TCTA, TCTC, TCTG and TCTT). The individual position HLAB_272 is a four allelic polymorphism. 86 alleles carry an A at this position, 253 alleles a C, 81 alleles a G and 159 alleles a T. With this a resolution of 56.3 % for HLA-B*1302, HLA-B*1501, HLA-B*1801, HLA-B*4001, HLA-B*4402 and HLA-B*4403, 72.5 % for HLA-B*0801, HLA-B*3501, HLA-B*3503, HLA-B*5101 and HLA-B*5701, and 85.1 % for HLA-B*0702 can be obtained. By using the microhaplotype strategy these resolutions increase dramatically for some alleles. For the frequent alleles HLA-B*1302, HLA-B*1501, HLA-B*1801, HLA-B*4001, HLA-B*4402 and HLA-B*4403 the resolution is still 56.3 %. The alleles HLA-B*0801 and HLA-B*3501 can be unambiguously distinguished from 76.5 % of all other alleles. The frequent alleles HLA-B*3503, HLA-B*5101 and HLA-B*5701 can be distinguished from 96.2 % of the alleles.

Another interesting side-effect of using microhaplotypes is that with one marker individual HLA alleles can be identified. For example, for HLAB_272f1 the microhaplotype CCTT represents only the allele HLA-B*0805, and ACTG represents only HLA-B*1404.

By selecting the markers this way, the resolution of the individual frequent alleles is between 96.6 % and 100 % in HLA-A, and 98.6 % and 100 % in HLA-B. Resolution is listed in the table 35 and 36.

Frequent alleles	Resolution	Number of rare alleles which can not be distinguished from the frequent alleles
HLA-A*0101	99,4	2
HLA-A*0201	96,6	12
HLA-A*0301	98,9	4
HLA-A*2301	98,9	4
HLA-A*2402	96,8	11
HLA-A*2902	98,3	6
HLA-A*3001	99,7	1
HLA-A*3002	100,0	0

Table 35: Resolution of the individual frequent alleles of HLA-A with 16 markers

Frequent alleles	Resolution	Number of rare alleles which can not be distinguished from the frequent alleles
HLA-B*0702	99,0	6
HLA-B*1302	99,8	1
HLA-B*1501	98,6	9
HLA-B*1801	99,8	1
HLA-B*1801	99,8	1
HLA-B*3501	99,5	3
HLA-B*3503	99,8	1
HLA-B*4001	99,5	3
HLA-B*4402	99,4	4
HLA-B*4403	99,0	6
HLA-B*5101	98,9	7
HLA-B*5701	100,0	0

Table 36: Resolution of the individual frequent alleles of HLA-B with 18 markers

The figures 16 and 17 show the interrelationship between the number of markers and the resulting resolution for HLA-A and HLA-B respectively. For HLA-A 45 markers are needed to achieve a 100 % resolution and 53 markers for HLA-B, respectively. Resolution in this context indicates how many rare HLA alleles can be distinguished from the frequent HLA alleles.

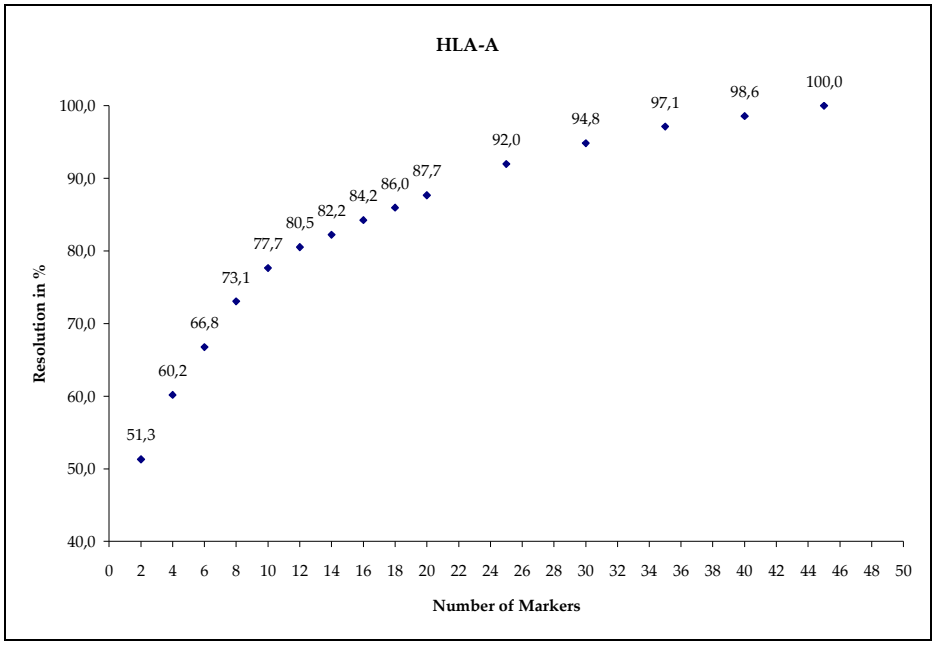


Figure 16: Resolution vs. number of markers (HLA-A)

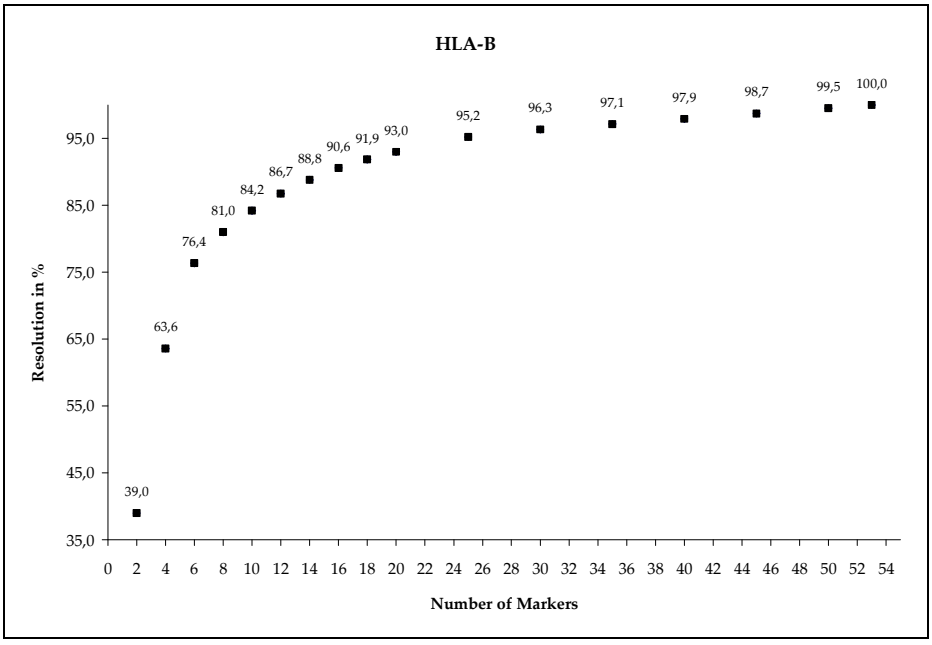


Figure 17: Resolution vs. number of markers (HLA-B)

Both graphs show similar characteristics, an asymptotic approximation to 100 % resolution. With just half of the total number of markers around 90 % resolution can be achieved, and as shown before, calculated for the individual frequent alleles the resolution is between 96 % and 100 %.
The markers which were selected for the HLA class I genes are listed table 37. They were selected based on the individual polymorphic positions, which are indicated in the names.

Markers HLA-A	**Marker HLA-B**
HLAA_81f1	HLAB_97f2
HLAA_98r2	HLAB_206f1
HLAA_123r2	HLAB_222r1
HLAA_241f1	HLAB_259f2
HLAA_259f2	HLAB_272f1
HLAA_268r2	HLAB_292f2
HLAA_282f1	HLAB_302f2
HLAA_299r2	HLAB_362f2
HLAA_355f2	HLAB_363r2
HLAA_413r1	HLAB_369f1
HLAA_453r1	HLAB_412f2
HLAA_502r2	HLAB_419f2
HLAA_527f1	HLAB_435r1
HLAA_539f1	HLAB_527f1
HLAA_559r1	HLAB_539f1
HLAA_571f2	HLAB_559r1
	HLAB_571r1
	HLAB_583r1

Table 37: Markers for HLA-A and HLA-B screening

The figures 18 and 19 show the consensus sequences of exons 2 and 3 of the HLA class I genes HLA-A and HLA-B. All polymorphic bases are coded following the rules of "Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences" recommended by the IUBMB. Further the markers used for the screening procedure are framed and indicated with an asterisk plus the number of position in the cDNA sequence. The black highlighted bases are the bases which are added during the extension reaction.

Figure 18: Consensus sequence of exon 2 and 3 of HLA-A. Polymorphic bases are coded by the IUBMB nomenclature. The microhaplotypes are framed and the numbers correspond to the position of the base in the complete cDNA sequence of the gene.

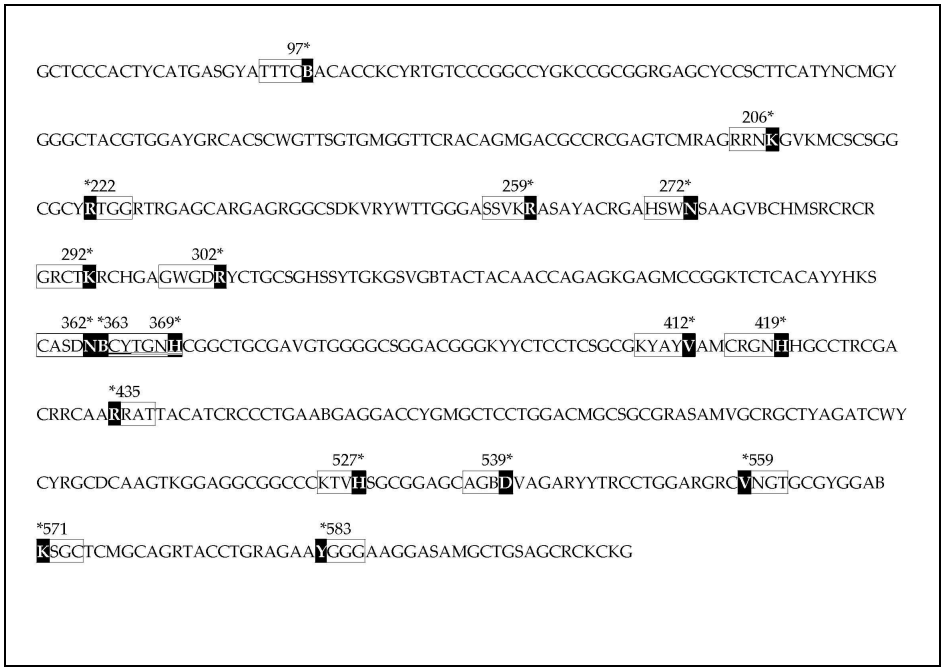


Figure 19: Consensus sequence of exon 2 and 3 of HLA-B. Polymorphic bases are coded by the IUBMB nomenclature. The microhaplotypes are framed and the numbers correspond to the position of the base in the complete cDNA sequence of the gene.

Figures 20 to 39 show the informativity of the different markers for the resolution of the individual frequent alleles. The resolution of the selected markers with

respect to the frequent HLA alleles is calculated here as percentage of all alleles (including the other frequent alleles) of HLA-A or HLA-B, respectively, which can be unambiguously distinguished from the individual frequent allele. The order of the markers on the X-axis is numerical. Figures 20 to 27 show the results for this calculation for HLA-A.

For the allele HLA-A*0101 the markers HLAA_268r2 and HLAA_539f1 are most important for the resolution. With either of these markers already 95,6 % of the other alleles can be distinguished from HLA-A*0101. Whereas the marker HLAA_123r2 is comparatively unimportant. This marker is also unimportant for HLA-A*0201, HLA-A*0301, HLA-A2301, HLA-A*2402 and HLA-A*2902, but it is very important for the identification of the frequent alleles HLA-A*3001 and HLA-A*3002. It is interesting that marker HLAA_527f1 separates 99.1 % of all alleles from the allele HLA-A*3001. The figures also show that not just one marker is important for the individual alleles, but a few of the selected markers provide a high resolution. Further it is interesting to see that in some cases there is a difference in informativity in terms of separation between exon 2 and exon 3. Exon 2 stretches from base 74 to 343, and exon 3 from base 344 to 619. For HLA-A*0101 markers in exon 3 provide more resolution than exon 2. For HLA-A*2301 and HLA-A*2402 the end of exon 2 and the very beginning of exon 3 are interesting. Exon 2 is informative for HLA-A*3001 and HLA-A*3002.

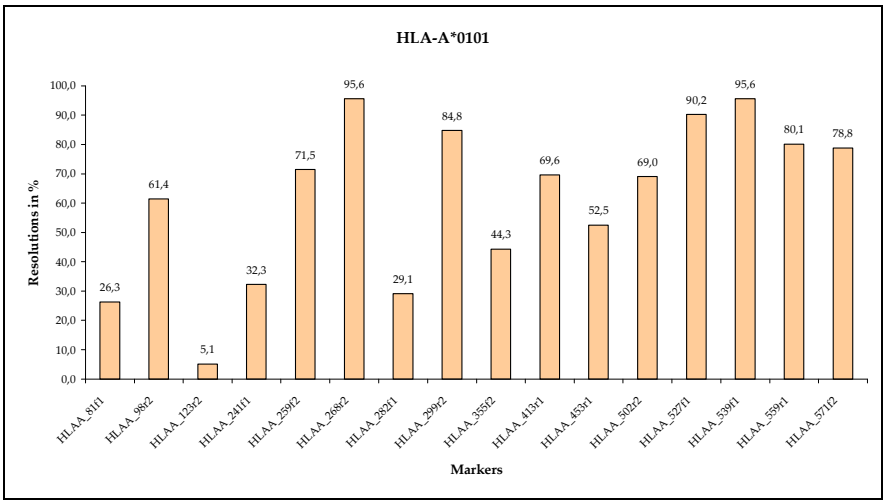


Figure 20

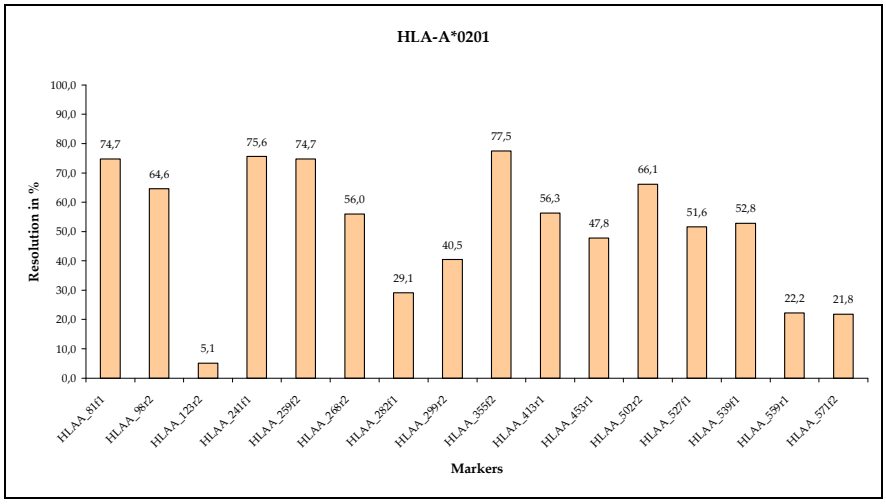


Figure 21

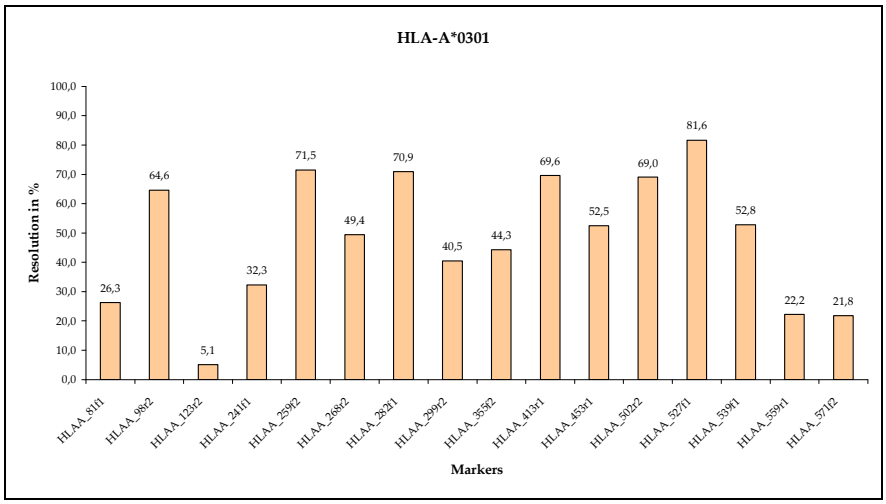


Figure 22

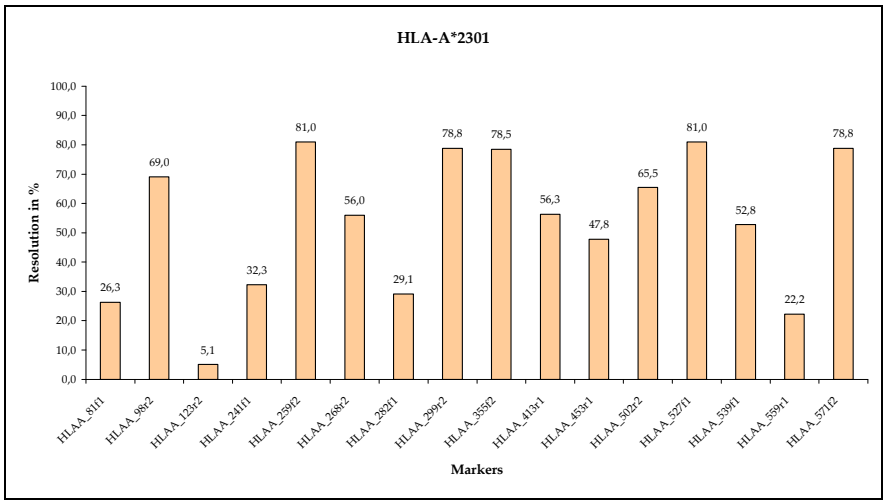


Figure 23

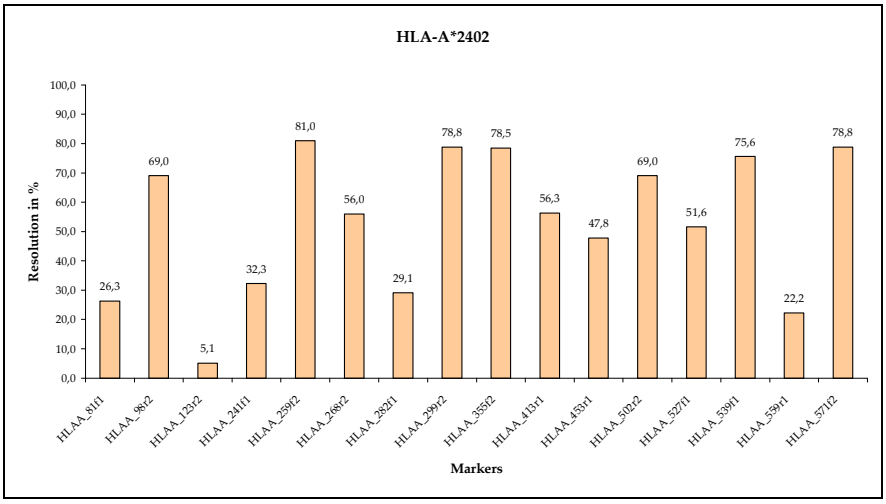


Figure 24

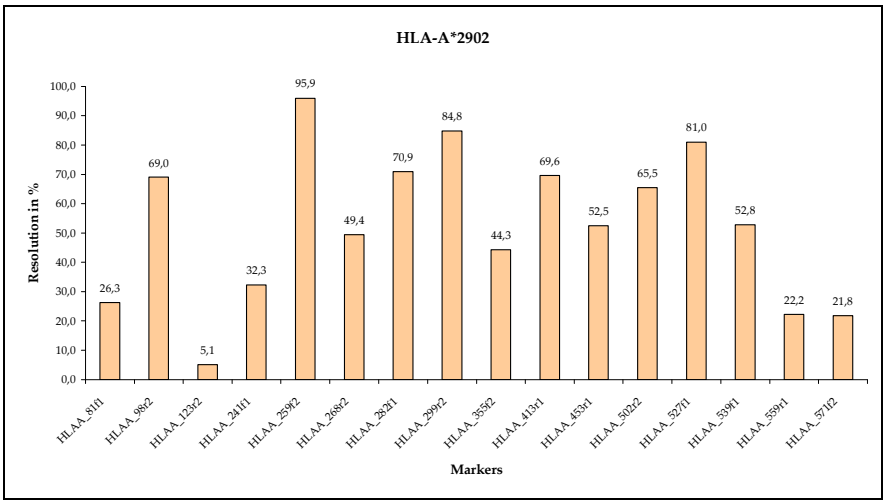


Figure 25

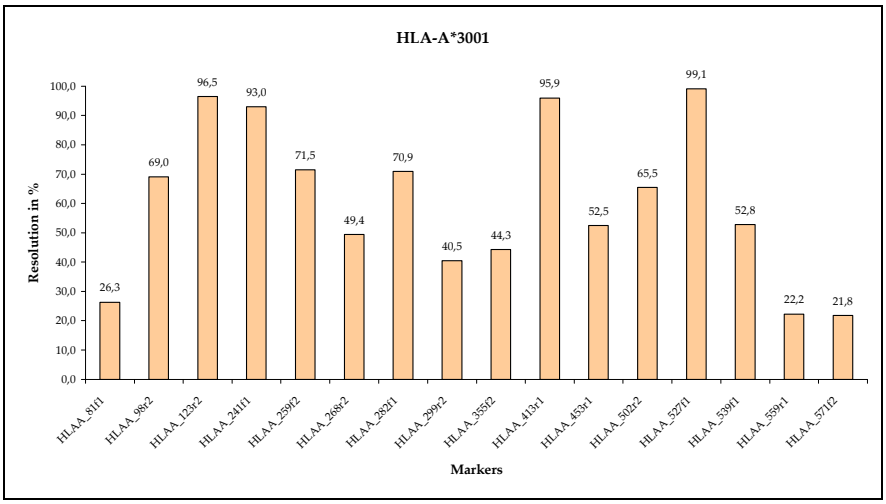


Figure 26

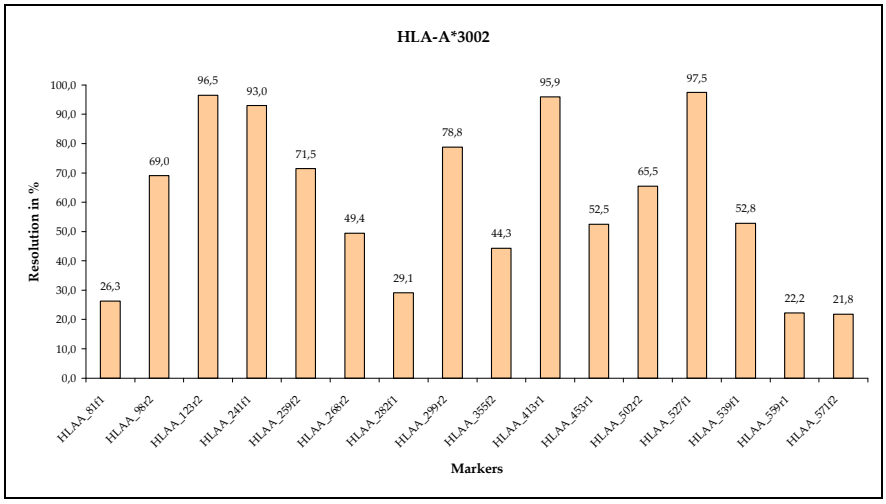


Figure 27

Figures 28 to 39 show the resolution of the markers for the individual frequent alleles of HLA-B. The characteristic is very similar to HLA-A. Here also each allele has its specific set of informative markers. For HLA-B the blocks of specific markers mostly do not spread over both exons. Here the blocks are smaller than for HLA-A. For example, the beginning of exon 3 (HLAB_369f1, HLAB_412f2 and HLAB_419f2) is very informative for the alleles HLA-B*1501, HLA-B*1801, HLA-B*5301, HLA-B*3503, HLA-B*4001, HLA-B*4402 and HLA-B*4403. For allele HLA-A*5701 exon 2 is very informative and exon 3 almost not at all. For the distinction of frequent and rare alleles, the informative markers are spread evenly across exon 2 and 3.

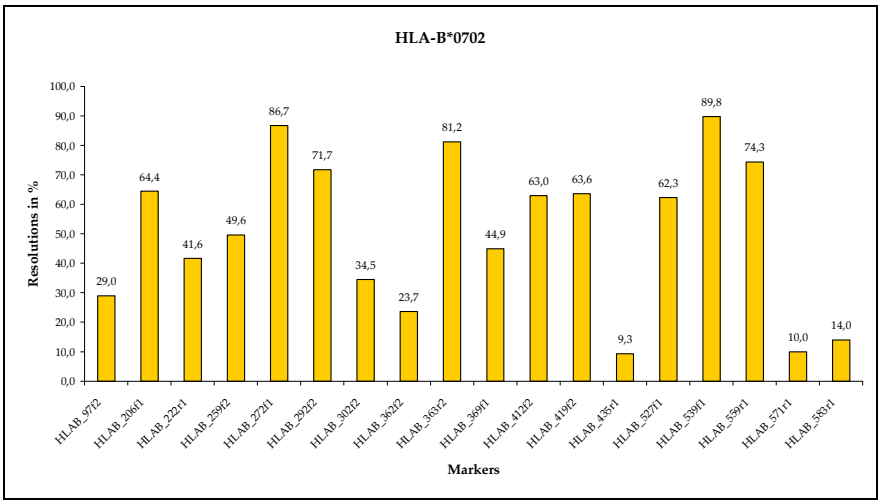


Figure 28

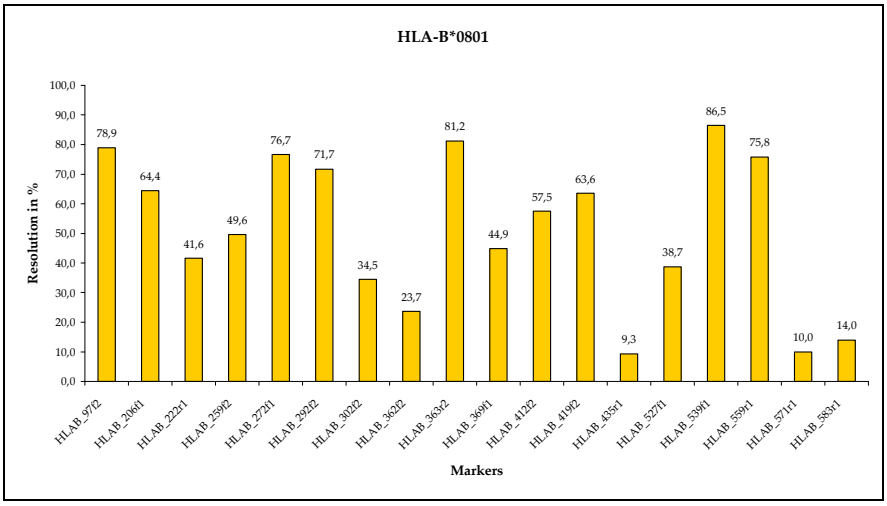


Figure 29

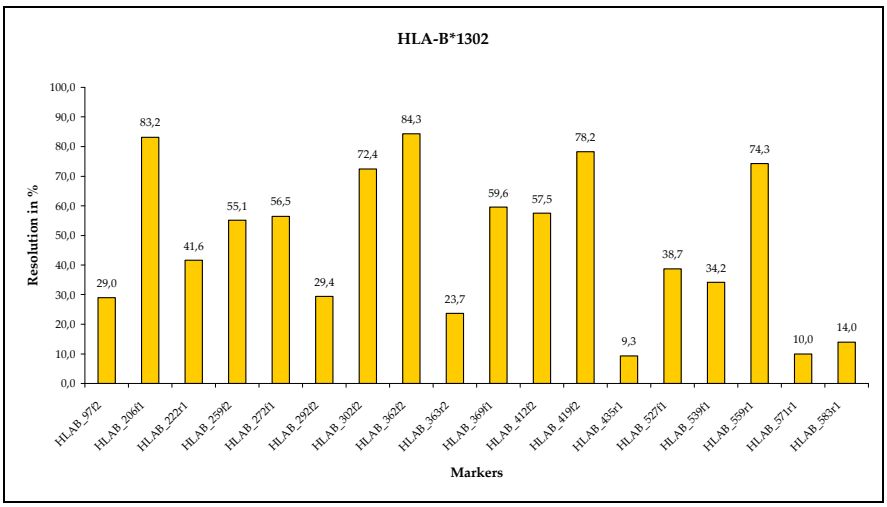


Figure 30

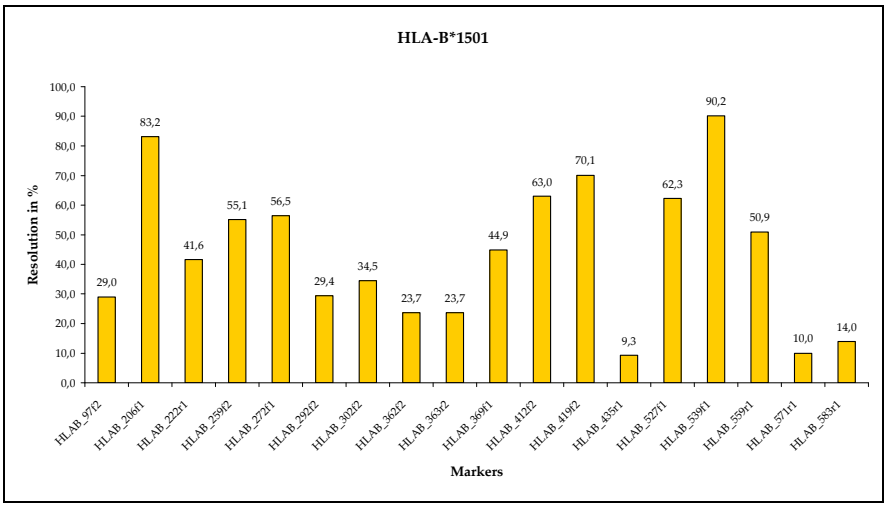


Figure 31

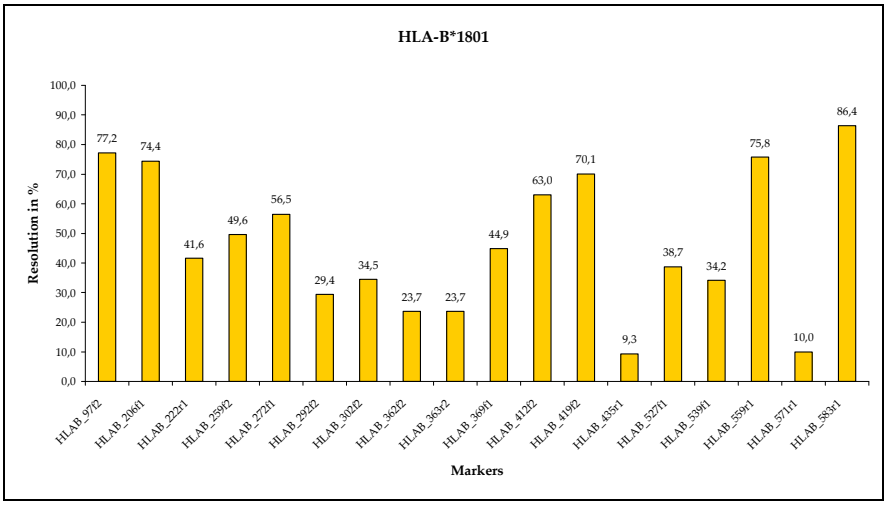


Figure 32

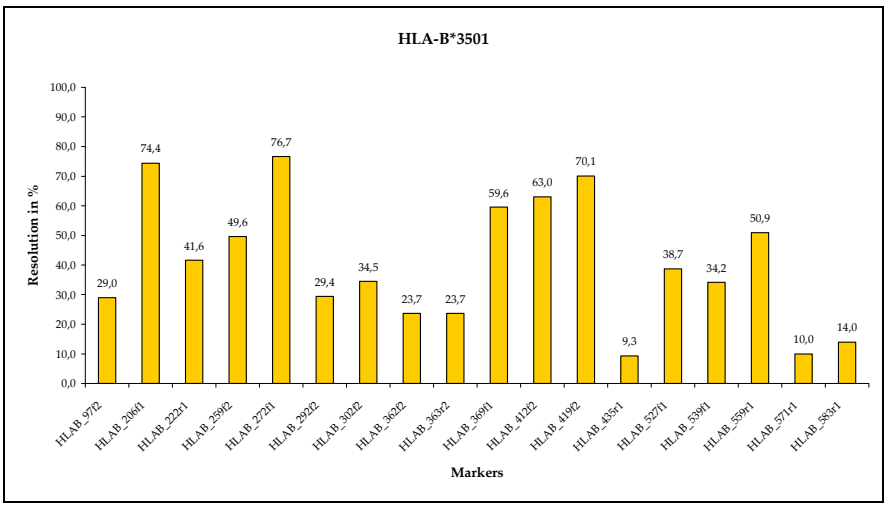


Figure 33

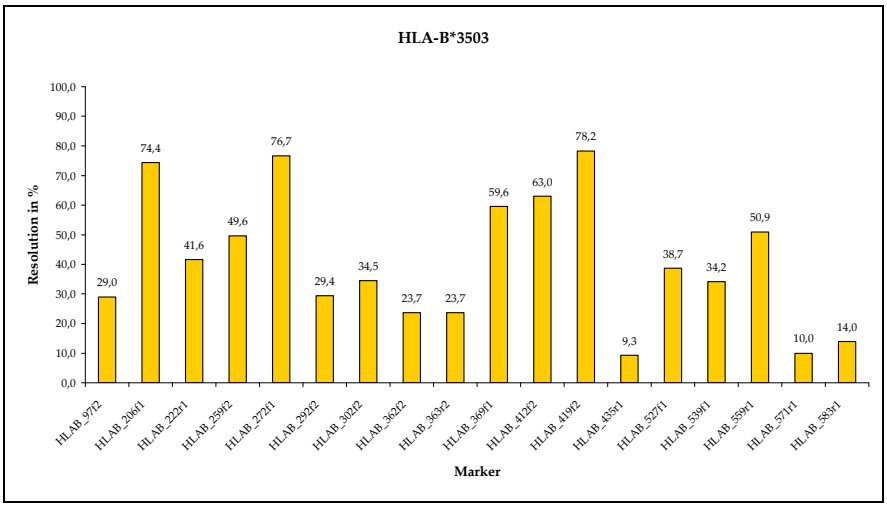


Figure 34

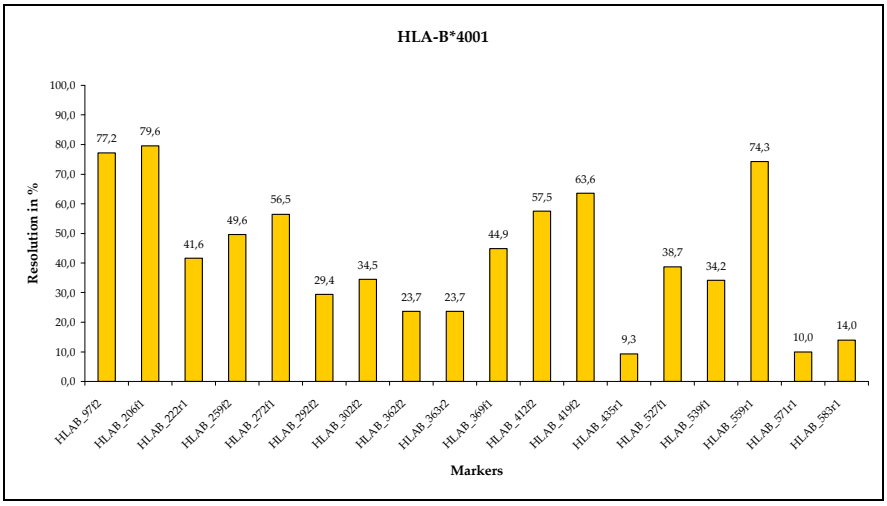


Figure 35

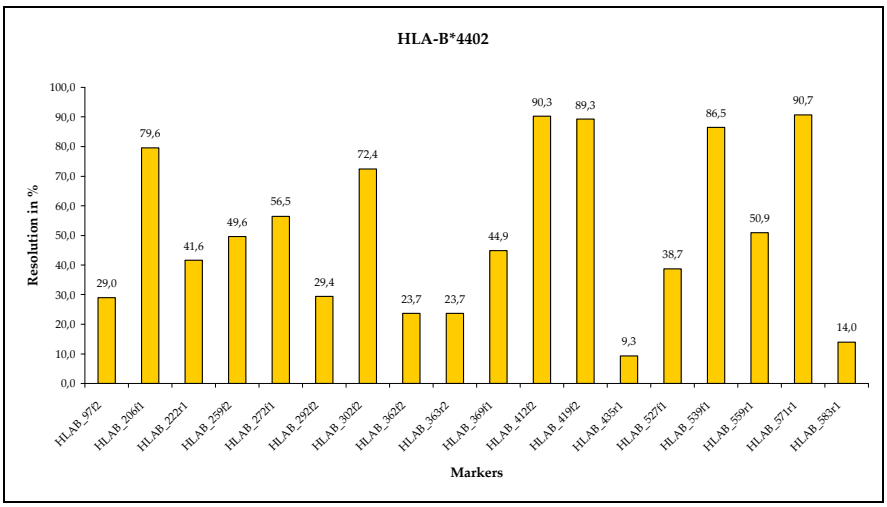


Figure 36

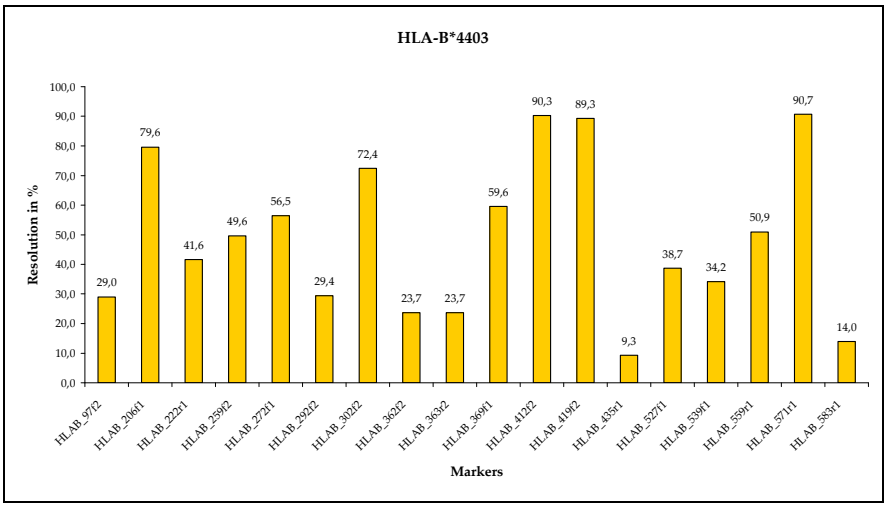


Figure 37

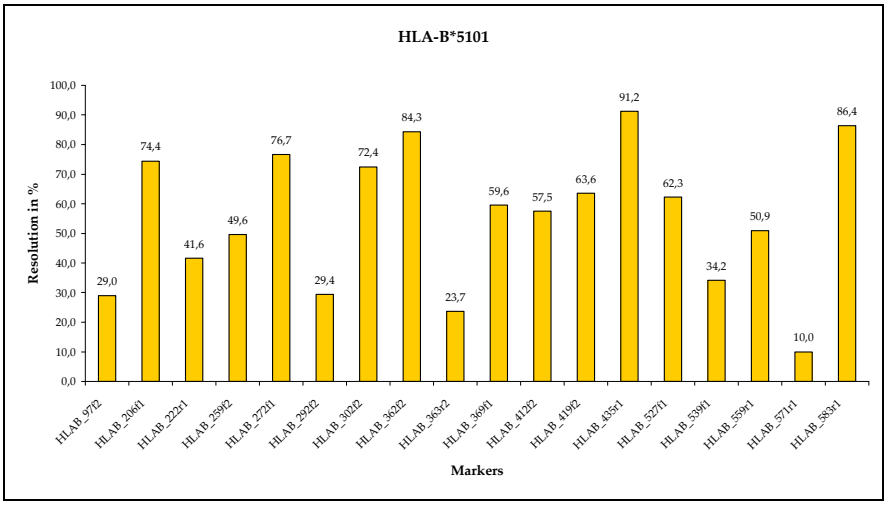


Figure 38

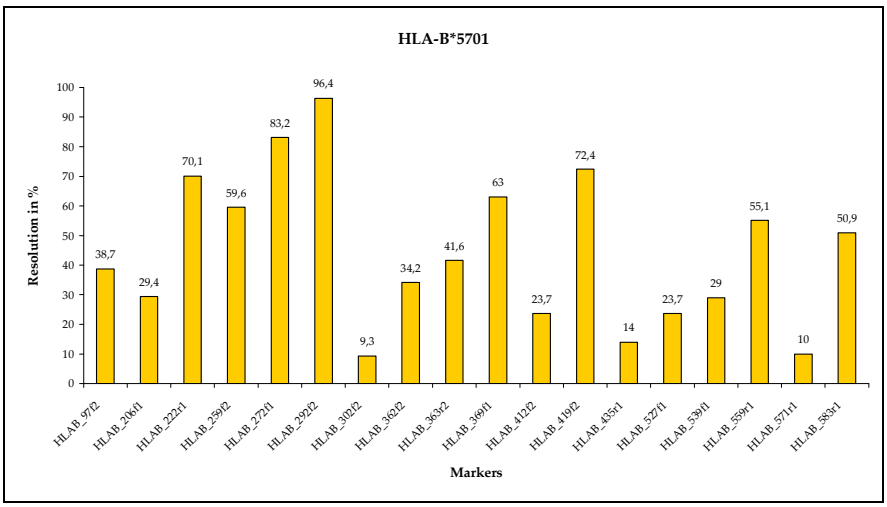


Figure 39

3.2.2 Marker selection and description for screening of the alleles of the gene HLA-DRB1

To select the markers for HLA-DRB1 the microhaplotype strategy was used. All single polymorphic bases were transformed into microhaplotypes and coded into a numeric code by the algorithm outlined in the methods, section 2.14.2. The markers for the HLA-DRB1 screening were than selected with the selection algorithm outlined in section 2.14.3. The criteria for the number of markers were based on the balance of informativity and costs.

Figure 40 shows the interrelationship between the number of markers and the resulting resolution for HLA-DRB1. The calculation of resolution was made with the optimal set of markers. An optimal set of markers contains the markers which

allow distinguishing a maximum number of rare alleles from frequent alleles for the permitted number of markers. A set of ten markers must not inevitably include the set which was obtained for five markers. The presented resolutions are for the separation of two groups of alleles (group1 = frequent alleles and group 2 = rare alleles).

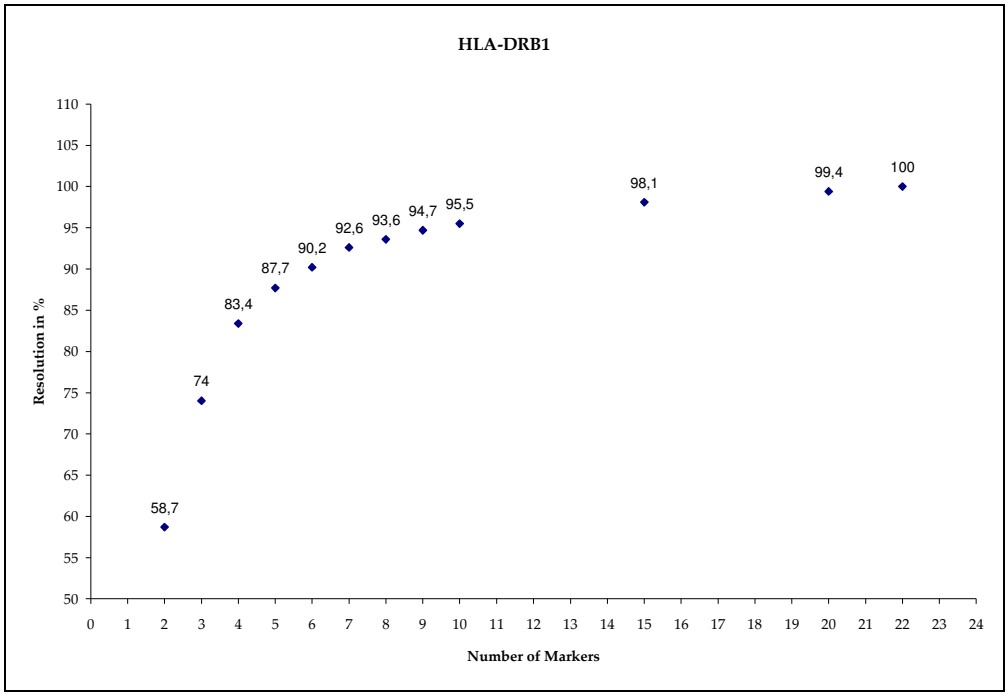


Figure 40: Resolution vs. Number of markers (HLA-DRB1)

The graph shows an asymptotic approximation to 100 % resolution. For a 100 % separation between frequent and rare HLA-DRB1 alleles a set of 22 microhaplotypes would have to be genotyped. These markers are DRB1_125f1, DRB1_165f1, DRB1_177f1, DRB1_184f1, DRB1_239f1, DRB1_262f1, DRB1_364f1, DRB1_115r1, DRB1_150r1, DRB1_188r1, DRB1_195r1, DRB1_196r1, DRB1_203r1, DRB1_220r1, DRB1_227r1, DRB1_262r1, DRB1_283r1, DRB1_295r1, DRB1_298r1, DRB1_304r1, DRB1_316r1 and DRB1_341r1.

With ten markers already 95.5 % of all rare alleles can be distinguished unambiguously from the frequent alleles, and the gain with every additional marker is small. If the calculation is made separately for each frequent allele the resolution lies between 98.94 % and 99.79 % (see table 38). In this case resolution

means that n % of all of the other alleles (including the other frequent alleles) can be distinguished unambiguously from the individual allele.

Frequent alleles	Resolution	Number of rare alleles which can not be distinguished from the frequent alleles
HLA_DRB1*0101	99,15	4
HLA-DRB1*0301	99,79	1
HLA-DRB1*0401	99,57	2
HLA-DRB1*0701	99,36	3
HLA-DRB1*1101	99,79	1
HLA-DRB1*1104	99,79	1
HLA-DRB1*1302	98,94	5
HLA-DRB1*1501	99,15	4

Table 38: Resolution of the individual frequent alleles of HLA-DRB1 with ten markers

The ten markers which give this resolution are HLADR_122r2, DRB_125r1, DRB_196f1, DRB_197r1, DRB_227f1, DRB_261r1, DRB_286f1, DRB_299f1, DRB_308r1 and DRB_341f1.

Of these ten markers originally only DRB_125r1 to DRB_341f1 (9 markers) were selected for screening. Since co-amplification of the other HLA-DRB genes is one of the most fatal flaws for DNA-based methods of HLA typing, an additional marker was included. Marker HLADR_122r2 genotypes a set of polymorphic bases, which allows a specific identification of HLA-DRB genes. At this position eleven different microhaplotypes could be observed. The microhaplotypes AATAT, AACAT, AACCT, AGAGG, AGTTT, CGTCT, CGGGT, GGGGT and TGGGT represent unambiguously one of the HLA-DRB1 alleles. AGTAT represents all alleles of the HLA-DRB5 gene and some HLA-DRB1*07 alleles. AGTCT represents all alleles of HLA-DRB3 and HLA-DRB7, plus the alleles HLA-DRB1*1130 and HLA-DRB1*1446. Finally, AGTGT corresponds to all alleles of HLA-DRB4 and HLA-DRB6, and AACAC the gene HLA-DRB9, respectively. By including this marker the specificity of the PCR amplification can be controlled. All PCR primers for HLA-DRB1 were particularly designed for the different HLA-DRB1 alleles. Since the homology between the different HLA-DRB genes is very high (around 90 % to 95 %), these PCR primers could anneal at the other HLA-DRB genes as well. Specific cycling conditions are chosen to avoid

these situations. If the PCR amplification for the same DNA sample were carried out in separate reactions, slight variation in the conditions could cause imbalances in the amplification results. In one reaction no co-amplification of the HLA-DRB2 – 9 genes will be observed, in another reaction exactly this happened. To avoid this complication, the PCR for one DNA sample was carried out in a 50 µl volume. After EXO I/SAP treatment aliquots of the PCR were distributed for typing of the different markers. This way it is assured that each marker had the same template for the extension reaction, and amplification mistakes of the PCR are detectable for all markers. In the final analysis signals resulting from co-amplifications can be eliminated.

Figure 41 shows the consensus sequence of exon 2 of HLA-DRB1. All polymorphic bases were coded following the rules of "Nomenclature for Incompletely Specified Bases in Nucleic Acid Sequences" recommended by the International Union of Biochemistry and Molecular Biology (IUBMB). Exon 2 of HLA-DRB1 ranges from base 101 to 370 of the complete cDNA sequence. Ten markers were selected. The microhaplotypes are framed and additionally the bases which are added during the extension reaction are highlighted with a black square, and labelled with an asterisk and a number. The number corresponds to the position of the marked base in the cDNA sequence, and the asterisk indicates the corresponding base in the figure. In the cases that the asterisk is on the left side of the number, the microhaplotype is in reverse orientation. In cases where the asterisk is on the right side the microhaplotype is forward directed.

Figure 41: Consensus sequence of exon 2 of HLA-DRB1. Polymorphic bases are coded by the IUBMB nomenclature. Selected microhaplotypes are framed. Numbers correspond to the position of the base in the complete cDNA sequence of the gene.

Figures 42 to 49 display the informativity of the individual markers in relation to the individual frequent alleles. The figures show that for each frequent allele a set of markers is informative. The markers DR_122r2, DRB_125r1, DRB_197r1 and DRB_299f1 are of utmost importance for the alleles HLA-DRB1*0101. For the allele HLA-DRB1*0301 the marker DRB_308f1 separates 90.6 % of the other alleles. Even 98.7 % of the other alleles can be distinguished from the allele HLA-DRB1*0701 by marker DRB_125r1. The lowest informativity for all frequent alleles is given by marker DRB_341f1. This marker has only a fine tuning character for the resolution.

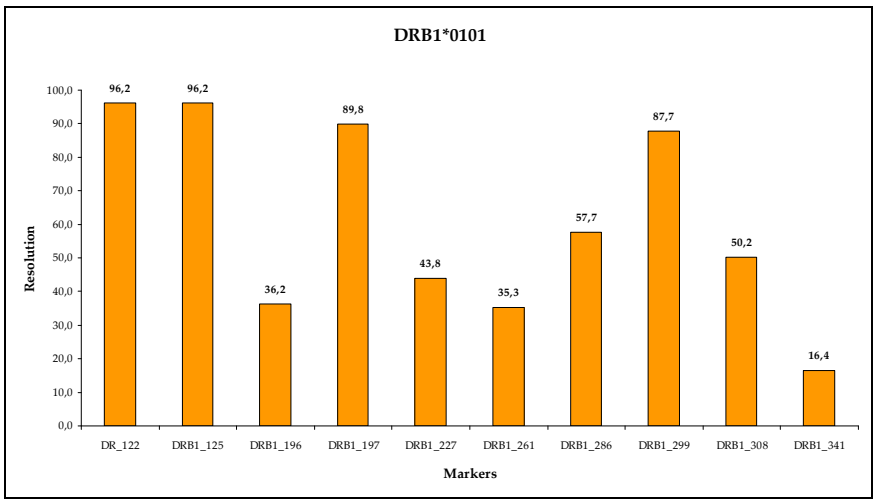


Figure 42

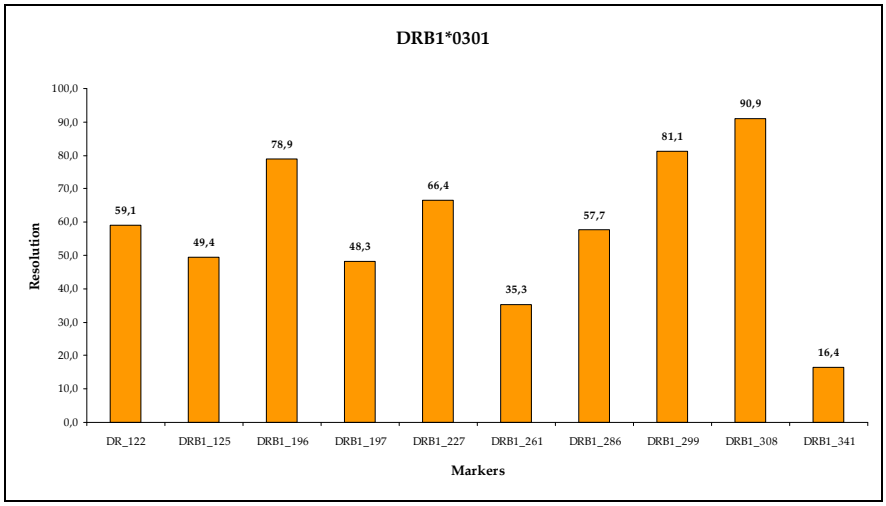


Figure 43

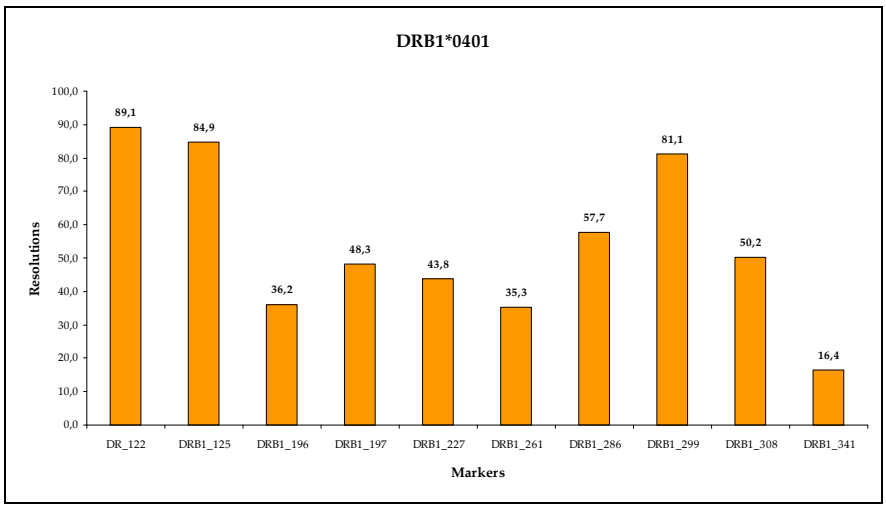


Figure 44

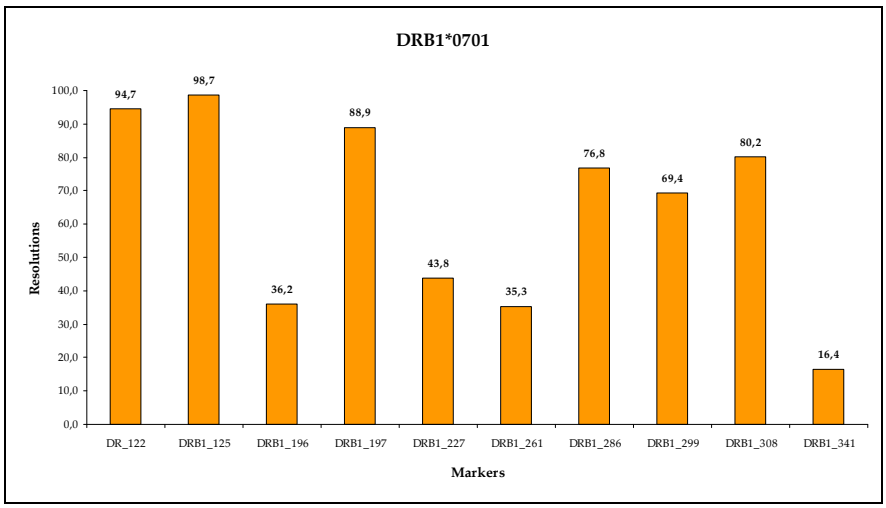


Figure 45

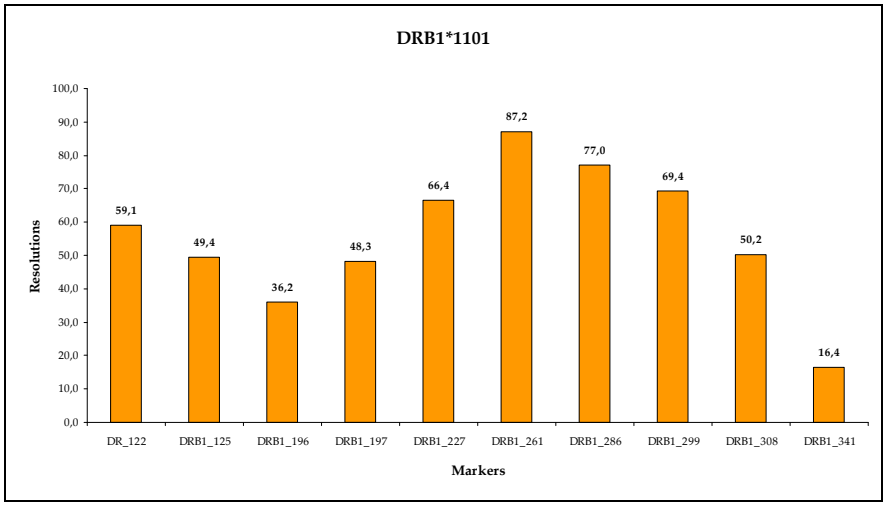


Figure 46

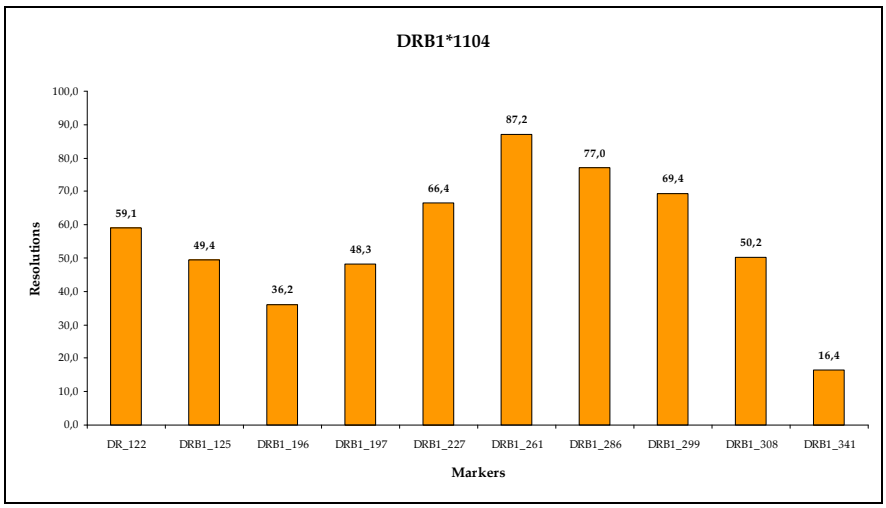


Figure 47

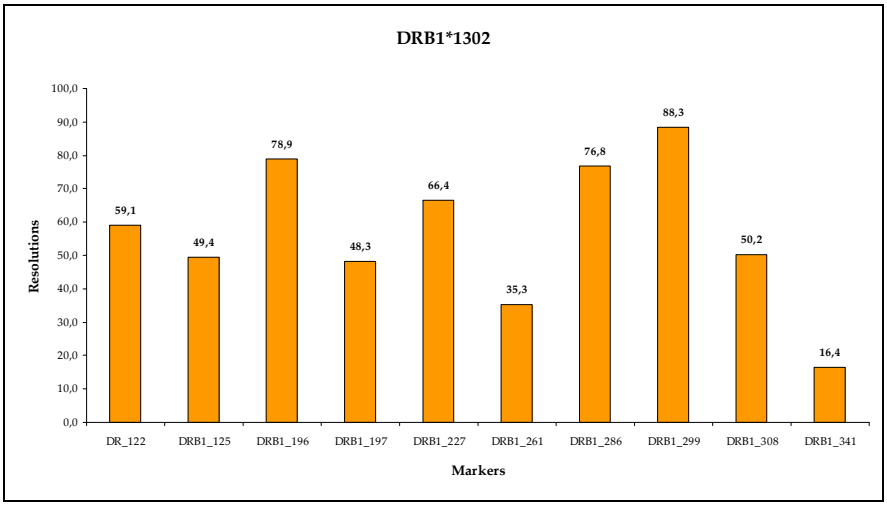


Figure 48

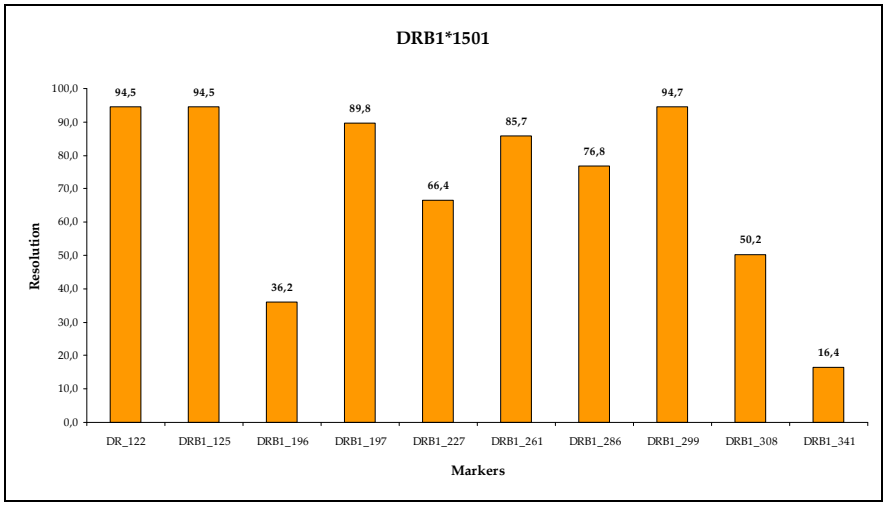


Figure 49

3.3 GOOD Assay for HLA Screening

The GOOD Assay is a method for SNP genotyping by single base primer extension and detection by mass spectrometry. The assay has been extensively described in section 1.6 "The GOOD Assay – a potential method for HLA Screening".

Usually one extension primer is used for SNP genotyping. In the human genome usually SNPs are bi-allelic, and found on average every 500 to 1000 bases[110]. However, the situation is very different in the MHC region. The polymorphic bases are very close to each other and often have more than two alleles.

As previously shown, the high density of polymorphism is in this case an advantage. Hybridization-based assays allow detecting the haplotype of polymorphisms because they reveal the sequence of this fragment. Cross hybridizations and high background can cause problems with these assays. Therefore an internal control such as an extension reaction can help resolve these problems and add a level of specificity. Only if the extension primer perfectly matches the template is an extension possible. The correct sequence can than be detected by the extension product. In the GOOD assay the primer is reduced to a core sequence of four to five bases. These are the added base of the extension reaction and three or four bases of the 3'end of the primer.

For each marker a pool of primers was designed. The combination of primers contains slightly varying sequences so that all sequences of the HLA alleles are

[110] Brooks, A.J.; 1999, Gene, 234 (2): 177 - 186

accommodated by a perfectly matching primer. The pool of primers guarantees that at least one primer is perfectly matched. The added base together with the base composition of the core sequence of the primer identifies a microhaplotype. Each microhaplotype in the pool has an unique mass by which it is recognised in the mass spectrum. In some cases a unique mass for a primer in a pool can not be achieved only by base composition. For example the pool for marker HLAA_292 contains five primers. Primers HLAA_2924_2f20 and HLAA_2925_2f20 have identical core sequences. Due to this these two primers express the same microhaplotypes (GACTC and GACTG). In the same pool the primer HLAA_2921_2f20 represents the microhaplotype GAGTC. This microhaplotype and GACTG have identical masses and can not be distinguished by mass spectrometry. However by using two different charge tags on the primers a mass shift of 14 Da is created. HLAA_2921_2f20 was charge tagged with CT -14 and HLAA_2924_2f20 and HLAA_2925_2f20 with CT 0. As a result GAGTC (CT -14) is represented by the mass 1801.6 Da and GACTG (CT 0) by the mass 1815.7 Da.

3.3.1 Evaluation of molecular biology

3.3.1.1 Generic PCR amplifications

For genotyping by the GOOD assay a PCR has to be carried out. The resulting product is used as the template for the extension reaction. Therefore, the PCR is of utmost importance. Since the MALDI-MS is very sensitive to salts (Na+, K+) and detergents it is better to avoid these substances. Good primer design is important to obtain effective and accurate PCR results. Annealing temperatures of the primers have to match well. The primer design options are extremely limited in the MHC region, because most of the genes are copies of each other (e.g. HLA-A, HLA-B and HLA-C have a homology of around 90%). Consequently the set of PCR primers outlined in table 2 and 3 were used. Most of these PCR primers were designed by other HLA research groups and verified by the International Histocompatibility Working Groups. The primers Bx1-2_f22, Bx1-3_f22 and BINT3-2_r23 were designed and added to the system in the context of this project. These additional primers are required to assure the amplification of all known HLA-B alleles.

The primer system for the HLA-DRB1 fragment is made up of nine forward primers and one reverse primer (see table 2). Some of the forward primers are likely to amplify one of the other HLA-DRB genes. Due to this it is of importance to use very stringent conditions. For all forward primers the optimal annealing temperature was measured individually, and than by comparison of the results the optimal annealing temperature for the complete system was defined. Variation of parameters in the polymerase chain reaction such as annealing temperature, magnesium concentration and the use of different DNA polymerases such as Platinum Taq, Platinum Taq High Fidelity and AmpliTaq Gold were examinated. Optimised parameters are Platinum Taq High Fidelity as DNA polymerase, a buffer with magnesium concentration between 1.5 mM (HLA-A) and 1.7 mM (HLA-B and HLA-DRB1) and annealing temperatures of 61°C for HLA-DRB1, 68.5°C for HLA-B and 70°C HLA-A. Optimised conditions of the three PCRs are detailed in section ("2.4 Generic PCR Amplification of Target Loci").

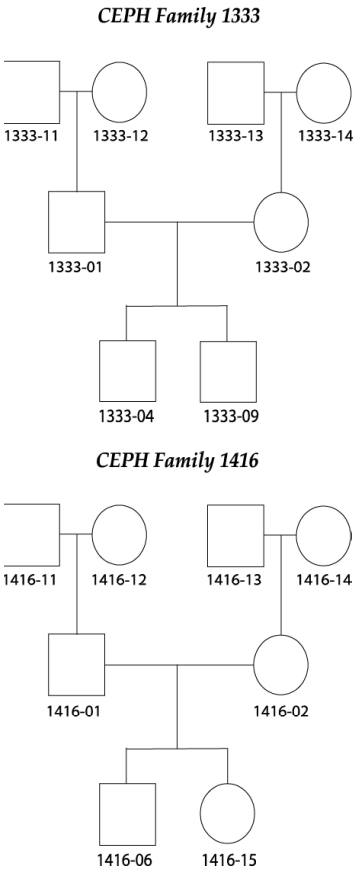
3.3.1.2 Extension reactions

The extension reaction is the key step of the GOOD assay. For the extension reaction several thermocycling profiles with different reaction parameters such as variation of units of different enzymes (ThermoSequenase, TMA31FS and ThermiPol), magnesium and ⍟-S-ddNTP concentrations were tested. TMA31FS is a newly engineered DNA polymerase, which preferentially incorporates ddNTPs over dNTPs[111], thus makes the removal of excess dNTPs from the PCR in principle redundant. However, as the inclusion of the SAP digestion step leads to improved signal-to-noise ratios the digest was routinely carried out, which eliminates the advantage of the TMA31FS for the purpose of this work. ThermiPol and ThermoSequenase is in principle the same enzyme. The main advantage, which finally made ThermiPol the enzyme of choice, is the extremely low price compared to other enzymes such as ThermoSequenase.
The respective terminating ⍟-S-ddNTPs were titrated against each other to investigate any preferential incorporation. Total concentration did not exceed 0.5 mM, since higher concentrations significantly disturb the enzyme.

[111] Sauer, S. et al.; 2002, Nucleic Acid Research, 30: e22

To determine the best temperature profile several profiles were tested. The inclusion of temperature ramping was an optimal solution. A slow ramping of 0.3°C/sec from the melting temperature down to the annealing temperature gives a competive advantage to the fully matching extension primers, since the melting temperature of fully matching DNA double strands is higher then the temperature of a DNA double strand that includes a mismatch. The optimised conditions of the extension reactions for all markers are outlined in section "2.7 Extension Reactions".

3.4 Microhaplotype-based Pre-screening of CEPH Families



The method for HLA type pre-screening, which has been developed here, will be exemplarily demonstrated for HLA-DRB1 on two families of the CEPH panel, family 1333 and family 1416. The HLA types of the members of these families are known and were kindly provided by Dr. Howard Cann (Foundation Jean Dausset - CEPH). The pedigrees are displayed in figure 50. The use of CEPH families has the advantage that for some individuals the HLA types are well studied and further the pedigree helps to control the correctness of the individual microhaplotype results, since the results have to follow Mendel's law of heritance.

Figure 50: Pedigree of two CEPH families (from http://locus.umdnj.edu/nigms/ ceph/ceph.html)

3.4.1 Pre-screening of HLA-DRB1

In the following the results for the HLA-DRB1 locus of three individuals of the CEPH family 1416 will be presented. This is exemplary for all other loci and individuals. The members of the family 1416 are 1416-11 (Paternal Grandfather;

A), 1416-12 (Paternal Grandmother; B) and 1416-01 (Father; C) as the offspring. The spectra are listed by the markers. In the spectra the masses, respectively the microhaplotypes, which are possible for the corresponding marker, are indicated by coloured bars. In table 39 the microhaplotypes and their colour codes, for the consecutively shown spectra, are presented. The colours are chosen randomly by the software "Helixir" and every time when "Helixir" opens spectra for one of the markers, other colours are selected.

DRB1_125r1	DRB1_196f1	DRB1_197r1	DRB1_227f1	DRB1_261r1
GGGA	GAAT	ACGT	ACTA	GGGG
GTGA	GAGC	TCGT	AGTT	GGAG
ATGA	GAGT	CCGT	AGTA	TGAG
TTGA	GAGA	ACGC		CGAG
CTGA	GAGG	ACTT		GGAC
ATAA		ACCT		CCAG
		TCCT		

DRB1_286f1	DRB1_299f1	DRB1_308r1	DRB1_341f1	HLADR_122r2
GACC	ACAA	GGGT	GGGC	GGGGT
AACA	ACAG	AGGT	GGGT	TGGGT
GACT	ACGA	TGGT		CGGGT
GACA	AGGC	CGGT		AGTGT
	AGAA	GCGT		AGTAT
	AGAG			AGTTT
	GGAG			AATAT
				AGTCT
				AACAT
				AGAGG
				CGTCT
				AACCT

Table 39: Possible microhaplotypes of HLA-DRB1 markers and their colour code in the "Helixir" picture.

Below are the spectra of marker DRB1_125r1 (figure 51) of three individuals presented. In the spectrum of the grandfather (A) two peaks appear, one at the mass of the microhaplotype GGGA (the left peak) and another one at the mass of the microhaplotype ATAA. Both are well identifiable by the colour bars (red and green). In the spectrum of the grandmother (B) also two peaks are visible. The one on the left appears at the mass of the microhaplotype ATGA and the one on the right at the microhaplotype CTGA. In the spectrum of the offspring, the transmission of the parental microhaplotypes is observed. The microhaplotype CTGA is the maternal one and ATAA is the paternal microhaplotype. This

principle can be followed up through the spectra of all markers. This procedure was used for the quality control of the method.

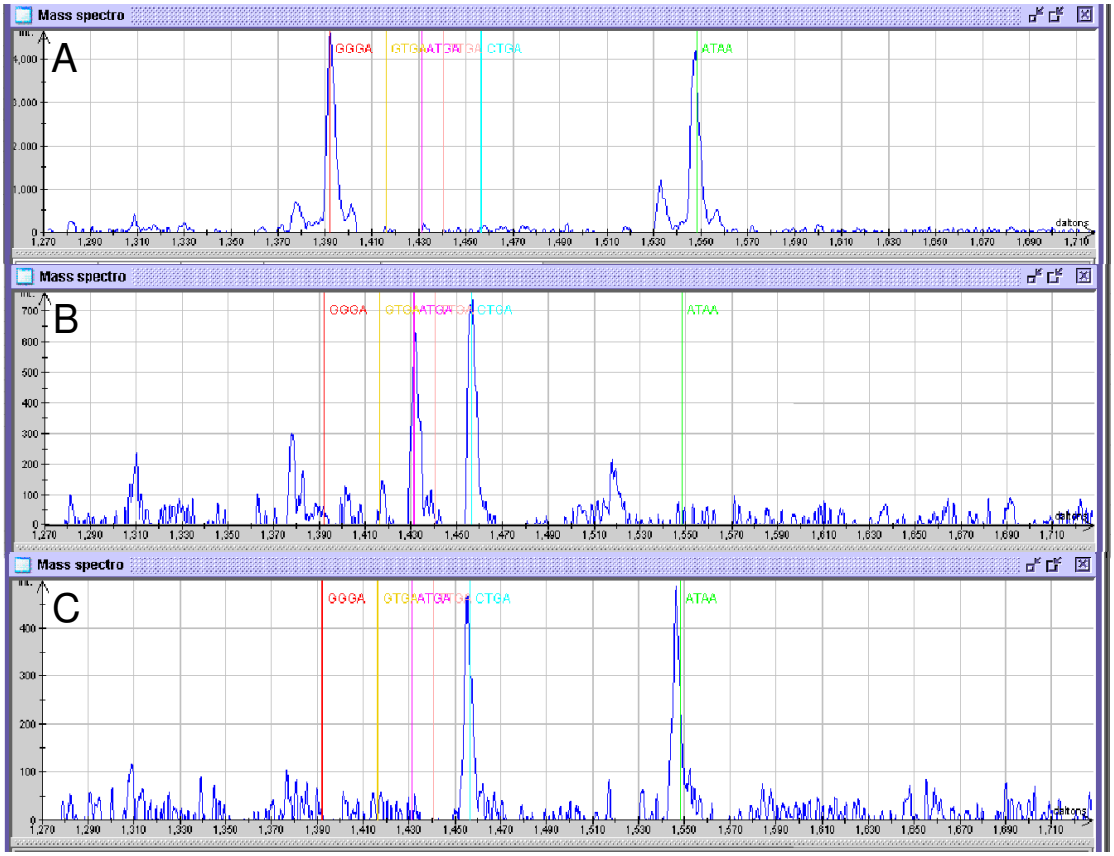


Figure 51: Spectra of the marker DRB1_125r1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

The figures 52 to 60 show the spectra of the markers DRB1_196f1 to HLADR_122r2.

Figure 52: Spectra of the marker DRB1_196f1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

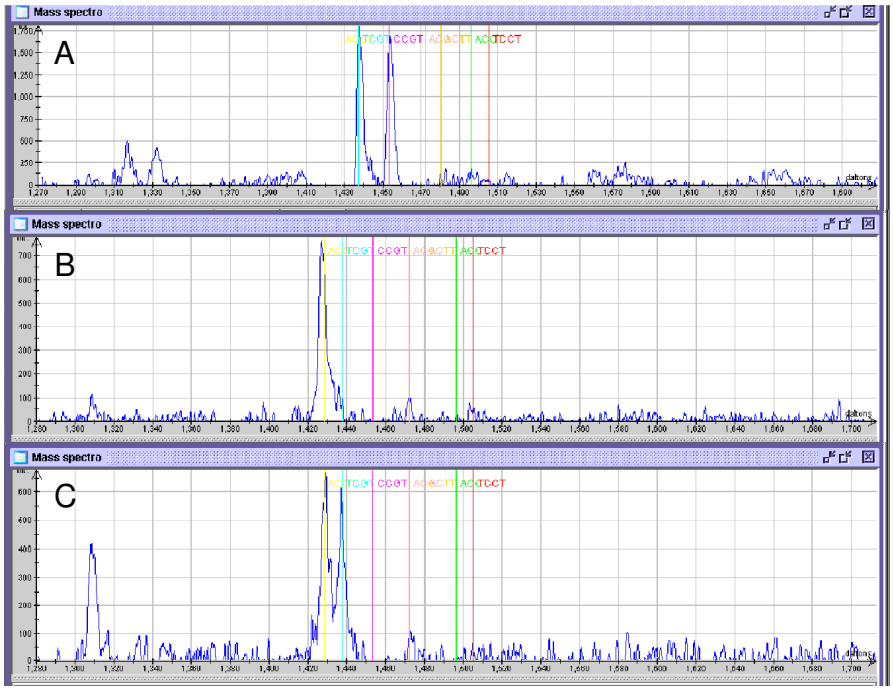


Figure 53: Spectra of the marker DRB1_197r1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

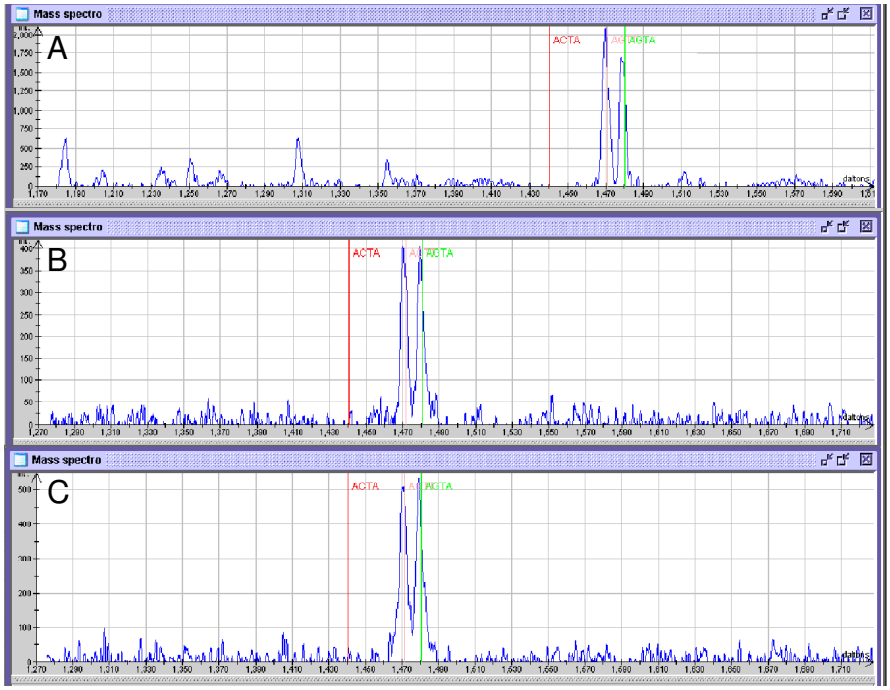


Figure 54: Spectra of the marker DRB1_227f1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

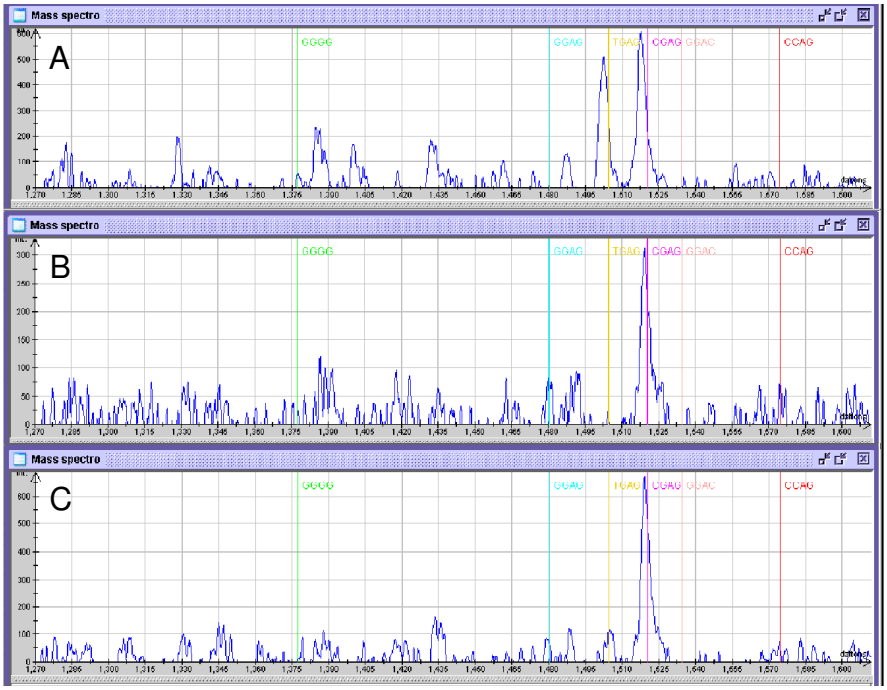


Figure 55: Spectra of the marker DRB1_261r1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

Figure 56 Spectra of the marker DRB1_286f1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

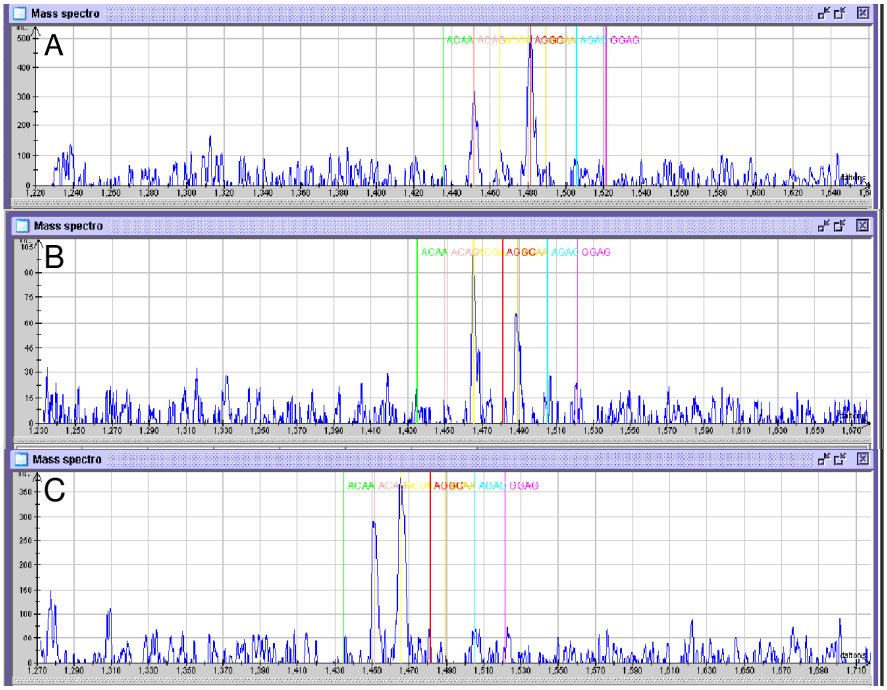


Figure 57: Spectra of the marker DRB1_299f1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

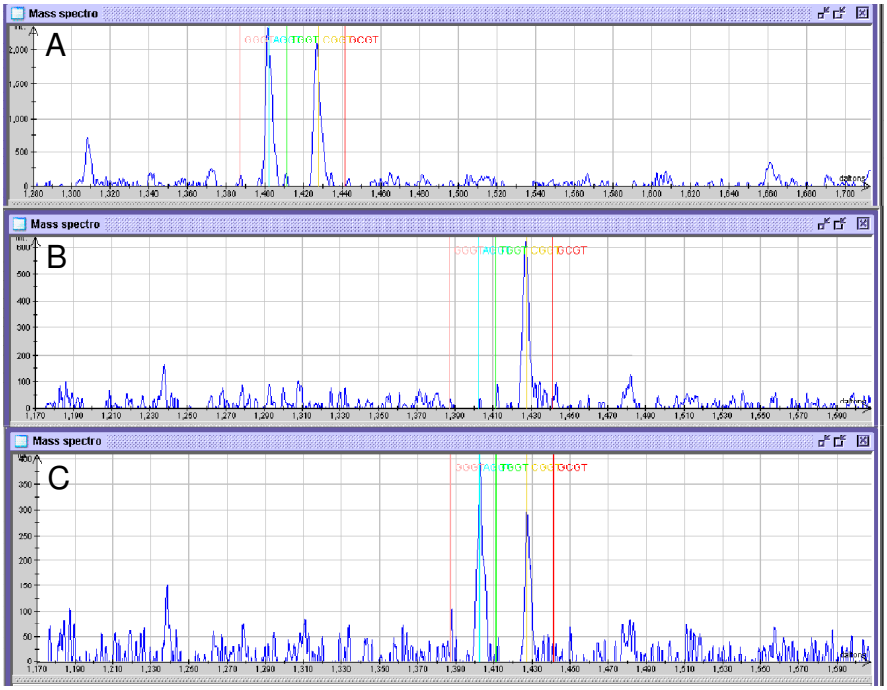


Figure 58: Spectra of the marker DRB1_308r1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

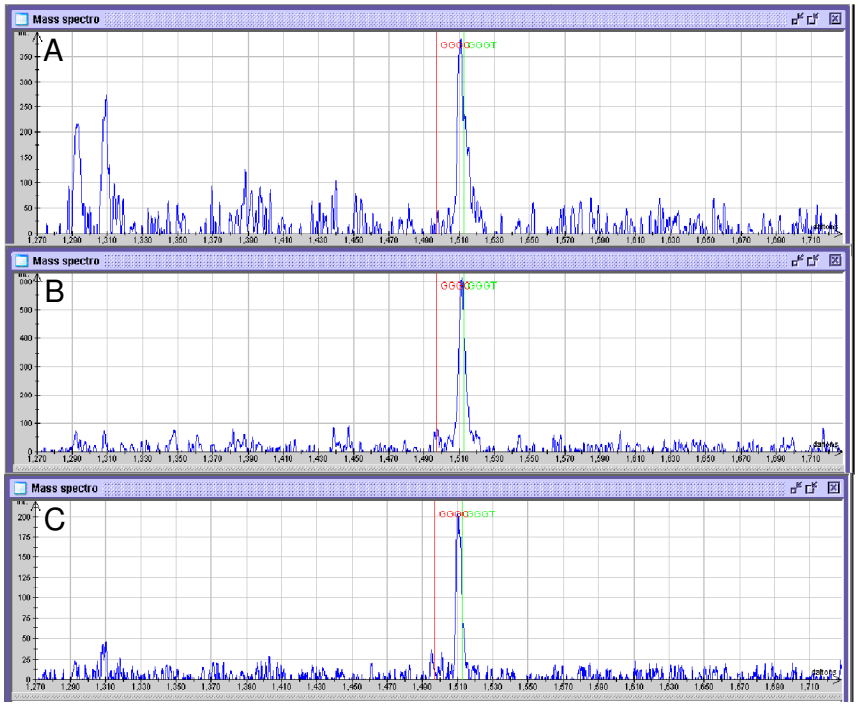


Figure 59: Spectra of the marker DRB1_341f1 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

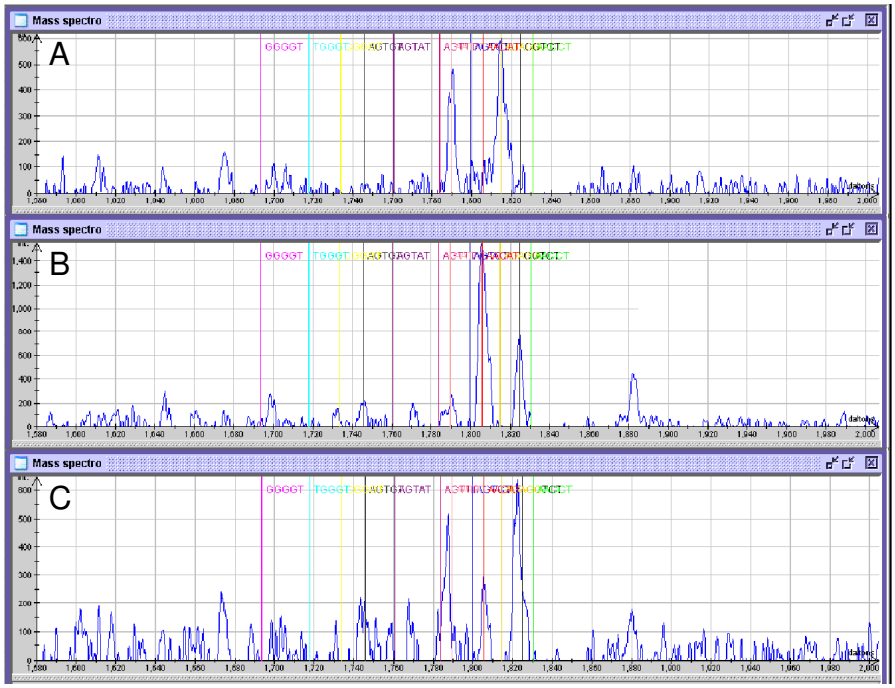


Figure 60: Spectra of the marker HLADR_122r2 of three individuals 1416-11 (A), 1416-12 (B) and 1416-01 (C)

Table 40 shows the determined microhaplotypes of the members of the two families of the CEPH library. In this table the individuals who are presented with spectra are highlighted grey.

Individual	HLADR_122r2	DRB1_125r1	DRB1_196f1	DRB1_197r1	DRB1_227f1
1333-01	CGGGT; CGTCT	GTGA; CTGA	GAGT; GAGT	ACGT; TCGT	AGTA; AGTT
1333-02	CGGGT; CGTCT	GTGA; CTGA	GAGT; GAGT	ACGT; ACGT	AGTA; AGTA
1333-04	CGGGT; CGTCT	GTGA; CTGA	GAGT; GAGT	ACGT; ACGT	AGTA; AGTA
1333-09	CGGGT; CGTCT	GTGA; CTGA	GAGT; GAGT	ACGT; TCGT	AGTA; AGTT
1333-11	CGGGT; CGTCT	CTGA; GTGA	GAGT; GAGT	ACGT; ACGT	AGTT; AGTA
1333-12	CGTCT; AGAGG	CTGA; GGGA	GAGT; GAGT	TCGT; CCGT	AGTT; AGTA
1333-13	CGTCT; AGAGG	CTGA; GGGA	GAGT; GAGT	ACGT; CCGT	AGTA; AGTT
1333-14	CGGGT; AGTTT	GTGA; TTGA	GAGT; GAGT	ACGT; ACGC	AGTA; AGTA
1416-01	AATAT; CGTCT	ATAA; CTGA	GAGT; GAGA	TCGT; ACGT	AGTA; AGTT
1416-02	AACAT; CGTCT	ATGA; CTGA	GAGT; GAGA	ACGT; ACGT	AGTA; AGTT
1416-06	AACAT; CGTCT	ATGA; CTGA	GAGT; GAGA	ACGT; ACGT	AGTA; AGTT
1416-11	AATAT; AGAGG	ATAA; GGGA	GAGT; GAGT	TCGT; CCGT	AGTA; AGTT
1416-12	AACAT; CGTCT	ATGA; CTGA	GAGT; GAGA	ACGT; ACGT	AGTA; AGTT
1416-13	CGTCT; CGTCT	CTGA; CTGA	GAGT; GAGA	ACGT; ACGT	AGTT; AGTT
1416-14	AACAT; AATAT	ATGA; ATAA	GAGT; GAGT	ACGT; TCGT	AGTA; AGTA
1416-15	AATAT; CGTCT	ATAA; CTGA	GAGT; GAGA	TCGT; ACGT	AGTA; AGTT

Individual	DRB1_261r1	DRB1_286f1	DRB1_299f1	DRB1_308r1	DRB1_341f1
1333-01	CGAG; GGAG	GACT; GACC	ACAG; GGAG	TGGT; AGGT	GGGT; GGGT
1333-02	CGAG; CGAG	GACT; GACA	ACAG; ACAA	TGGT; CGGT	GGGT; GGGT
1333-04	CGAG; CGAG	GACT; GACA	ACAG; ACAA	TGGT; CGGT	GGGT; GGGT
1333-09	CGAG; GGAG	GACT; GACC	ACAG; GGAG	TGGT; AGGT	GGGT; GGGT
1333-11	GGAG; CGAG	GACT; GACT	ACGA; ACAG	CGGT; TGGT	GGGT; GGGT
1333-12	GGAG; TGAG	GACC; GACT	GGAG; ACAG	AGGT; CGGT	GGGT; GGGT
1333-13	CGAG; TGAG	GACA; GACA	ACAA; AGGC	CGGT; CGGT	GGGT; GGGT
1333-14	CGAG; CGAG	GACT; GACC	ACAG; GGAG	TGGT; CGGT	GGGT; GGGT
1416-01	CGAG; CGAG	GACA; GACA	ACAG; ACGA	AGGT; CGGT	GGGT; GGGT
1416-02	CGAG; CGAG	GACC; GACA	AGAA; ACGA	CGGT; CGGT	GGGT; GGGT
1416-06	CGAG; CGAG	GACC; GACA	AGAA; ACGA	CGGT; CGGT	GGGT; GGGT
1416-11	CGAG; TGAG	GACA; GACA	ACAG; AGGC	AGGT; CGGT	GGGT; GGGT
1416-12	CGAG; CGAG	GACC; GACA	AGAA; ACGA	CGGT; CGGT	GGGT; GGGT
1416-13	AGGA; CGAG	GACT; GACA	ACAG; ACGA	CGGT; CGGT	GGGT; GGGT
1416-14	CGAG; CGAG	GACC; GACA	AGAA; ACAG	CGGT; AGGT	GGGT; GGGT
1416-15	CGAG; CGAG	GACA; GACA	ACAG; ACGA	AGGT; CGGT	GGGT; GGGT

Table 40: Results of individuals of families 1333 and 1416 from the CEPH library, 1416-01, 1416-02 and 1416-06 are highlighted grey

These results can be aligned in different combinations. According to 2^n (n = heterozygote markers), for the individual 1416-01, with n = 7 heterozygote results of the 10 markers, 128 microhaplotype combinations could be created. Only for 10 microhaplotype combinations of these 128 combinations, corresponding to HLA-DRB1 alleles are known in the HLA allele database.

These are:

	HLADR_122r2	DRB1_125r1	DRB1_196f1	DRB1_197r1	DRB1_227f1	DRB1_261r1	DRB1_286f1	DRB1_299f1	DRB1_308r1	DRB1_341f1
1	AATAT	ATAA	GAGT	TCGT	AGTA	CGAG	GACA	ACAG	AGGT	GGGT
2	CGTCT	CTGA	GAGA	ACGT	AGTT	CGAG	GACA	ACGA	CGGT	GGGT
3	CGTCT	CTGA	GAGT	ACGT	AGTT	CGAG	GACA	ACGA	CGGT	GGGT
4	CGTCT	CTGA	GAGA	ACGT	AGTT	CGAG	GACA	ACAG	CGGT	GGGT
5	CGTCT	CTGA	GAGT	ACGT	AGTA	CGAG	GACA	ACAG	CGGT	GGGT
6	CGTCT	CTGA	GAGT	ACGT	AGTT	CGAG	GACA	ACAG	CGGT	GGGT
7	CGTCT	CTGA	GAGA	ACGT	AGTA	CGAG	GACA	ACGA	CGGT	GGGT
8	CGTCT	CTGA	GAGT	ACGT	AGTA	CGAG	GACA	ACGA	CGGT	GGGT
9	CGTCT	CTGA	GAGT	TCGT	AGTT	CGAG	GACA	ACGA	CGGT	GGGT
10	CGTCT	CTGA	GAGT	TCGT	AGTA	CGAG	GACA	ACGA	CGGT	GGGT

Table 41: Microhaplotype combinations based on the results of individual 1416-01, which can be found in the look- up table

There is strong linkage of the microhaplotypes AATAT (marker HLADR_122r2) and ATAA (DRB1_125r1), as well as of the microhaplotypes CGTCT (HLADR-122r2) and CTGA (DRB1_125r1). So far no allele has been identified that carries either the microhaplotype combination AATAT – CTGA or CGTCT – ATAA for these markers. However, the combination AATAT – ATAA is specific for the allele HLA-DRB1*070102 which is presented in the four-digit typing result as HLA-DRB1*0701.

According to the analysis procedure as described in section 2.14.5 the criteria for the identification of the alleles were, only allele combinations are possible which are heterozygous at each position that came up with a heterozygous result. Therefore, for example, an allele combination with the haplotype combination No. 2 and No. 3 from the table 41 is not possible, because in this case not all as heterozygous identified positions are also heterozygous in the alleles. Therefore the haplotype combination if No. 1 in table 41 is one of the parental HLA alleles only No. 2 can represent the second HLA allele, unless there is not an unknown allele present. The microhaplotype combination No. 1, as already mentioned, represents only HLA-DRB1*070102, and No. 2 represents the alleles HLA-DRB1*1301, HLA-DRB1*1302, HLA-DRB1*1315, HLA-DRB1*1316, HLA-DRB1*1327, HLA-DRB1*1328, HLA-DRB1*1331, HLA-DRB1*1335, HLA-DRB1*1339, HLA-DRB1*1341, HLA-DRB1*1351, HLA-DRB1*1359, HLA-DRB1*1361. Of these alleles HLA-DRB1*1301 and HLA-DRB1*1302 are the most likely alleles.
Table 42 shows the most likely HLA-DRB1 types for the analysed individuals of the families 1333 and 1416 of the CEPH panel. Most likely are allele combinations which summed up have a likelihood of 99 %. All other allele combinations together have a likelihood of less the 1 %. The highlighted HLA types are the HLA types which are carried by the individual (According to the CEPH database[112]). In most instances the most likely HLA-DRB1 type was the same as determined by CEPH. However in the family 1416 some of the right genotypes were the second ranked results.

[112] www.ceph.fr

Barcode	Allele 1	Allele 2		Likelihood
1333_01				
	DRB1*0801	DRB1*1401	r r	0.803195579473563
	DRB1*0804	DRB1*1401	r r	0.111140997470402
	DRB1*0802	DRB1*1401	r r	0.0408424868173644
	DRB1*0806	DRB1*1401	r r	0.0351912496578525
1333_02				
	DRB1*0801	DRB1*1303	r r	0.810784373675483
	DRB1*0804	DRB1*1303	r r	0.112191085616743
	DRB1*0802	DRB1*1303	r r	0.0412283769231774
	DRB1*0806	DRB1*1303	r r	0.0355237454511397
1333_04				
	DRB1*0801	DRB1*1303	r r	0.810784373675483
	DRB1*0804	DRB1*1303	r r	0.112191085616743
	DRB1*0802	DRB1*1303	r r	0.0412283769231774
	DRB1*0806	DRB1*1303	r r	0.0355237454511397
1333_09				
	DRB1*0801	DRB1*1401	r r	0.803195579473563
	DRB1*0804	DRB1*1401	r r	0.111140997470402
	DRB1*0802	DRB1*1401	r r	0.0408424868173644
	DRB1*0806	DRB1*1401	r r	0.0351912496578525
1333_11				
	DRB1*0801	DRB1*1103	r r	0.810759077235515
	DRB1*0804	DRB1*1103	r r	0.112187585259367
	DRB1*0802	DRB1*1103	r r	0.0412270905994696
	DRB1*0806	DRB1*1103	r r	0.0355226371117051
1333_12				
	DRB1*1401	DRB1*1601	r r	0.990448543982072
1333_13				
	DRB1*1303	DRB1*1501	r f	0.934921868919902
	DRB1*1303	DRB1*1502	r r	0.0629035641590103
1333_14				
	DRB1*0801	DRB1*1001	r r	0.810890418006865
	DRB1*0804	DRB1*1001	r r	0.112205759343874
	DRB1*0802	DRB1*1001	r r	0.0412337692763193
	DRB1*0806	DRB1*1001	r r	0.035528391682563
1416_01				
	DRB1*0701	DRB1*1301	f r	0.605336767288411
	DRB1*0701	DRB1*1302	f f	0.394524341132929
1416_02				
	DRB1*0401	DRB1*1301	f r	0.605306247173283
	DRB1*0401	DRB1*1302	f f	0.394504449844372
1416_06				
	DRB1*0401	DRB1*1301	f r	0.605306247173283
	DRB1*0401	DRB1*1302	f f	0.394504449844372
1416_11				
	DRB1*0701	DRB1*1501	f f	0.935024214347444
	DRB1*0701	DRB1*1502	f r	0.0629104501805932
1416_12				
	DRB1*0401	DRB1*1301	f r	0.605306247173283
	DRB1*0401	DRB1*1302	f f	0.394504449844372

1416_13				
	DRB1*1101	DRB1*1301	f r	0.350425915683998
	DRB1*1104	DRB1*1301	f r	0.252700642999574
	DRB1*1101	DRB1*1302	f f	0.22838783462704
	DRB1*1104	DRB1*1302	f f	0.164696017276238
1416_14				
	DRB1*0401	DRB1*0701	f f	0.99989942764735
1416_15				
	DRB1*0701	DRB1*1301	f r	0.605336767288411
	DRB1*0701	DRB1*1302	f f	0.394524341132929

Table 42: Analysis output for the determination of the HLA-DRB1 types

3.4.2 Pre-screening of HLA-A and HLA-B types

Identical to the procedure of microhaplotype-based pre-screening of HLA-DRB1 types, the sample DNAs of the members of the families 1333 and 1416 from the CEPH panel were also pre-screened for their HLA-A and HLA-B types. The determined microhaplotypes of the different markers and samples are listed in the tables 43 to 46.

Based on these microhaplotyping results, the in table 48 shown HLA types have been estimated. The results are in complete agreement with the HLA types provided by the CEPH database.

	HLAA_81f1	HLAA_98r2	HLAA_123r2	HLAA_241f1	HLAA_259f2	HLAA_268r2	HLAA_282f1	HLAA_299r2
1333-01	CACC	TCACA; CCACA	CGGGG	GGGC	CCAGG; CGAGG	AATGT; AAAGT	CCAG; CCAC	TGGAC; AGAAC
1333-02	CACC	TCACA	CGGGG	GGGC	CCAGG	AATAT; AATGT	CCAC; CCAG	CGAAC; TGGAC
1333-04	CACC	TCACA; CCACA	CGGGG	GGGC	CCAGG; CGAGG	AATAT; AAAGT	CCAC	CGAAC; AGAAC
1333-09	CACC	TCACA	CGGGG	GGGC	CCAGG	AATGT	CCAG	TGGAC
1333-11	TCAC; CCAC	TCACA; CCACA	CGGGG	GGTC; GGGC	CGGGG; CGAGG	AAAGT	CCAC	TGGAC; AGAAC
1333-12	CACC	TCACA; CCACA	CGGGG	GGGC	CCAGG; CGAGG	AATGT; AAAGT	CCAG; CCAC	TGGAC; AGAAC
1333-13	CCAC; TCAC	TCACA	CGGGG	GGGC; GGTC	CCAGG; CGGGG	AATAT; AAAGT	CCAC	CGAAC; TGGAC
1333-14	TCAC; CCAC	TCACA	CGGGG	GGTC; GGGC	CGGGG; CCAGG	AAAGT; AATGT	CCAC; CCAG	TGGAC
	HLAA_355f2	**HLAA_413r1**	**HLAA_453r1**	**HLAA_502r2**	**HLAA_527f2**	**HLAA_539f1**	**HLAA_559r1**	**HLAA_571f2**
1333-01	CACCA; CACCC	GGCA; ACCA	CGAG; AGAG	AAGCG	CATGA; CATGT	AGTT; AGCA	ACGT	GGAGT; GGACG
1333-02	CACCA	GGCA	CGAG	AAGCG	CATGC; CATGA	AGCG; AGTT	CGGT; ACGT	GGACG; GGAGT
1333-04	CACCA; CACCC	GGCA; ACCA	CGAG; AGAG	AAGCG	CATGC; CATGT	AGCG; AGCA	CGGT; ACGT	GGACG
1333-09	CACCA	GGCA	CGAG	AAGCG	CATGA	AGTT	ACGT	GGAGT
1333-11	CACCG; CACCC	ACCA	AGAG	AAGCA; AAGCG	CATGT	AGTT; AGCA	ACGT	GGAGT; GGACG
1333-12	CACCA; CACCC	GGCA; ACCA	CGAG; AGAG	AAGCG	CATGA; CATGT	AGTT; AGCA	ACGT	GGAGT; GGACG
1333-13	CACCA; CACCG	GGCA; ACCA	CGAG; AGAG	AAGCG; AAGCA	CATGC; CATGT	AGCG; AGTT	CGGT; ACGT	GGACG; GGAGT
1333-14	CACCG; CACCA	ACCA; GGCA	AGAG; CGAG	AAGCA; AAGCG	CATGT; CATGA	AGTT	ACGT	GGAGT

Table 43: Microhaplotyping results of HLA-A of the family 1333 from the CEPH panel

	HLAB_97f2	HLAB_206f1	HLAB_222r1	HLAB_259f2	HLAB_272f1	HLAB_292f2
1333 01	TTTCT	GGAC; GAGA	ATGG; GTGG	CCGGA	TCTT; TCTG	GACTT; GACTG
1333 02	TTTCT	GAGA; GGAC	GTGG; ATGG	CCGGA; CGGGG	TCTA; ACAT	GACTG; GACTT
1333 04	TTTCT	GAGA; GGAC	GTGG; ATGG	CCGGA; CGGGG	TCTG; ACAT	GACTG; GACTT
1333 09	TTTCT	GAGA; GGAC	GTGG; ATGG	CCGGA	TCTA; TCTT	GACTG; GACTT
1333 11	TTTCT	GGAT; GAGA	ATGG; GTGG	CCGGG; CCGGA	TCTC; TCTG	GACTT; GACTG
1333 12	TTTCT	GGAC	ATGG	CCGGA	TCTT	GACTT
1333 13	TTTCT	GGAC; GAGA	ATGG; GTGG	CGGGG; CCGGA	ACAT; TCTA	GACTT; GACTG
1333 14	TTTCT; TTTCC	GAGA; GGAC	GTGG	CCGGA; CCGGG	TCTA; TCTC	GACTG; GACTT
	HLAB_302f2	**HLAB_362f2**	**HLAB_363r2**	**HLAB_369f1**	**HLAB_412f2**	**HLAB_419f2**
1333 01	GAGAG	CAGAG; CAGAC	GATGT	TGGC; CGGC	GCATG; GCATA	CAGTC; CAGTT
1333 02	GAGAG; GAGAA	CAGAG	CATGT; GATGT	CGGC; TGGC	GCATG	CAGTA; CAGTC
1333 04	GAGAG; GAGAA	CAGAC; CAGAG	GATGT	CGGC; TGGC	GCATA; GCATG	CAGTT; CAGTC
1333 09	GAGAG	CAGAG	CATGT; GATGT	CGGC; TGGC	GCATG	CAGTA; CAGTC
1333 11	GAGAG	CAGAG; CAGAC	GATGT	CGGC	GCATG; GCATA	CAGTC; CAGTT
1333 12	GAGAG; GAGAA	CAGAG; CAGAC	GATGT	TGGC	GCATG; GCATA	CAGTC; CAGTA
1333 13	GAGAA; GAGAG	CAGAG	GATGT; CATGT	CGGC; TGGC	GCATG	CAGTC; CAGTA
1333 14	GAGAG; GAGGA	CAGAG	CATGT; GATGT	CGGC; TGGC	GCATG; GTATA	CAGTA; CAGTT
	HLAB_435r1	**HLAB_527f1**	**HLAB_539f1**	**HLAB_559r1**	**HLAB_571r1**	**HLAB_583r1**
1333 01	GGAT	GTGT	AGCT	CTGT; ACGT	TGGC	TACC
1333 02	GGAT	GTGA; GTGT	AGCG; AGCT	GAGT; CTGT	TGGC	TACC
1333 04	GGAT	GTGT	AGCT	ACGT; CTGT	TGGC	TACC
1333 09	GGAT	GTGA; GTGT	AGCG; AGCT	GAGT; CTGT	TGGC	TACC
1333 11	GGAT	GTGA; GTGT	AGTG; AGCT	CTGT; ACGT	TGGC	TACC
1333 12	GGAT; AGAT	GTGT; GTGA	AGCT	CTGT	TGGC	TACC; CACC
1333 13	GGAT	GTGT; GTGA	AGCT; AGCG	CTGT; GAGT	TGGC	TACC
1333 14	GGAT	GTGA; GTGT	AGCG; AGGA	GAGT; ACGT	TGGC	TACC

Table 44: Microhaplotyping results of HLA-B of the family1333 from the CEPH panel

Based on the microhaplotyping results in tables 43 to 46 the HLA types of HLA-A and HLA-B were calculated. These types match the HLA types of these individuals provided by the CEPH-database.

Barcode	Allele 1	Allele 2		Likelihood
1333_01				
	A*0301	A*2402	f f	0.999679904727277
1333_02				
	A*0101	A*0301	f f	0.999930296597092
1333_04				
	A*0101	A*2402	f f	0.999715548627779
1333_09				
	A*0301	A*0301	f f	0.999947339881557
1333_11				
	A*0201	A*2402	f f	0.998330759255522
1333_12				
	A*0301	A*2402	f f	0.999679904727277
1333_13				
	A*0101	A*0201	f f	0.99858176442024
1333_14				
	A*0201	A*0301	f f	0.998545651694595
1416_01				
	A*0201	A*2301	f f	0.99828927171362
1416_02				
	A*0201	A*0205	f r	0.998329478576771
1416_06				
	A*0201	A*0201	f f	0.998597505573322
1416_11				
	A*2301	A*2301	f f	0.99969553135816
1416_12				
	A*0201	A*2501	f r	0.998593992608712
1416_13				
	A*0205	A*2501	r r	0.999999998617152
1416_14				
	A*0101	A*0201	f f	0.99858176442024
1416_15				
	A*0205	A*2301	r f	0.999692738450184

Table 45: Estimated HLA-A types of the families 1333 and 1416 from the CEPH panel

Barcode	Allele 1	Allele 2		Likelihood
1333-01				
	B*3501	B*3906	f r	0.999612278278644
1333-02				
	B*0702	B*5801	f r	0.999926103057903
1333-04				
	B*3906	B*5801	r r	0.999597261234961
1333-09				
	B*0702	B*3501	f f	0.999737406645754
1333-11				
	B*1501	B*3906	f r	0.999446641865479
1333-12				
	B*3501	B*5101	f f	0.989092365045609
	B*3501	B*5132	f r	0.0103444897390128
1333-13				
	B*0702	B*5801	f r	0.999926103057903
1333-14				
	B*0702	B*3701	f r	0.999799473566302
1416-01				
	B*4001	B*4101	f r	0.999194484931912
1416-02				
	B*1501	B*4901	f r	0.97715110811507
	B*1524	B*5001	r r	0.0224226988384785
1416-06				
	B*1501	B*4001	f f	0.999508617580113
1416-11				
	B*4101	B*4901	r r	0.999451854702427
1416-12				
	B*4001	B*4402	f f	0.99094272429437
1416-13				
	B*4402	B*4901	f r	0.99099569577341
1416-14				
	B*0801	B*1501	f f	0.999590700835219
1416-15				
	B*4101	B*4901	r r	0.999451854702427

Table 46: Estimated HLA-B types of families 1333 and 1416 from the CEPH panel. In the cases of individuals 1333-12 and 1416-02 respectively the first HLA-B types (high-lighted in grey) is right.

According to the identified HLA types of HLA-A, -B and -DRB1, and based on the family relations the HLA haplotypes can be established (table 49).

Family 1333

1333_01	11	12	A*2402-B*3906- DRB1*080[1426]	A*0301-B*3501- DRB1*1401
1333_02	13	14	A*0101-B*5801- DRB1*1303	A*0301-B*0702- DRB1*080[1426]
1333_04	1	2	A*2402-B*3906- DRB1*080[1426]	A*0101-B*5801- DRB1*1303
1333_09	1	2	A*0301-B*3501- DRB1*1401	A*0301-B*0702- DRB1*080[1426]
1333_11	0	0	A*2402-B*3906- DRB1*080[1426]	A*0101-B*3501- DRB1*1103
1333_12	0	0	A*0301-B*3501- DRB1*1401	A*2402-B*510[12]- DRB1*1601
1333_13	0	0	A*0101-B*5801- DRB1*1303	A*0201-B*5801- DRB1*1501
1333_14	0	0	A*0301-B*0702- DRB1*080[1426]	A*0201-B*3701- DRB1*1001

Family 1416

1416_01	11	12	A*2301-B*4101- DRB1*0701	A*0201-B*4001- DRB1*130[12]
1416_02	13	14	A*0205-B*4901- DRB1*130[12]	A*0201-B*1501- DRB1*0401
1416_06	1	2	A*0201-B*4001- DRB1*130[12]	A*0201-B*1501- DRB1*0401
1416_11	0	0	A*2301-B*4101- DRB1*0701	A*2301-B*4901- DRB1*150[12]
1416_12	0	0	A*0201-B*4001- DRB1*130[12]	A*2501-B*4402- DRB1*0401
1416_13	0	0	A*0205-B*4901- DRB1*130[12]	A*2501-B*4402 DRB1*110[14]
1416_14	0	0	A*0201-B*1501- DRB1*0401	A*0101-B*0801- DRB1*0701
1416_15	1	2	A*2301-B*4101- DRB1*0701	A*0205-B*4901- DRB1*130[12]

Table 47: HLA haplotypes of individuals from families 1333 and 1416 from the CEPH panel. Columns two and three indicate the parents of the individuals.

Notes to haplotypes in table 49:
In brackets the likelihoods of alleles based on their frequencies are given.
DRB1*080[1426] is a group of alleles with the following probabilities:
DRB1*0801 (0.81)
DRB1*0804 (0.12)
DRB1*0802 (0.04)
DRB1*0806 (0.03)

B*510[12] is a group of alleles with the following probabilities:
B*5101 (0.99)
B*5102 (0.01)

DRB1*130[12] is a group of alleles with the following probabilities:
DRB1*1301 (0.61)
DRB1*1302 (0.39)

DRB1*110[14] is a group of alleles with the following probabilities:
DRB1*1101 (0.58)
DRB1*1104 (0.42)

3.5 Microhaplotype-based Pre-screening at High throughput

After successful proof-of-principle using DNA samples of the CEPH panel, the method was performed on 655 individuals selected by the members of the MADO project, in a high-throughput format. All liquid handling was done using a BasePlate liquid handling robot. The accumulations of spectra were done automatically and unattended using the software provided with the Autoflex mass spectrometer. The microhaplotypes were determined automatically from the spectra using the "Helixir" software developed in the framework of the MADO project at the CNG. The software is described in detail in section 2.14.4 "Helixir" – reading raw data. The identification of the possible HLA types based on the determined microhaplotypes was done using the "HLAfamilies" software as well developed in the framework of the MADO project at the CNG.

For the markers HLADR_122r2 and DRB1_299f1 no results were obtained. Reasons for the failure of these markers are difficult to define. One reason could be that one or several of the components of the reaction were degraded. For instance, the ◎◎S-ddNTPs, which are used for the extension reaction, are quite sensitive. Also possible is a failure of the SAP digestion of the excess dNTPs from the PCR. In this case the polymerase of the extension reaction preferably incorporates dNTPs rather than the ◎-S-ddNTPs. This would lead to products which are not detectable with the used parameter set of the mass spectrometer. The other markers provided a return rate between 75 % (DRB1_261r1) and 98 % (DRB1_341f1). With this an estimation of HLA types could be provided for 97 % of the samples.

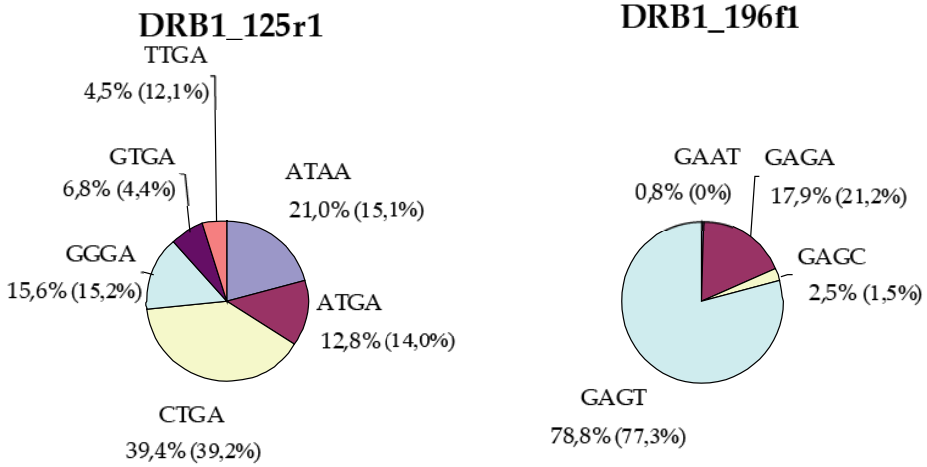


Figure 61: Microhaplotype frequencies found by genotyping of 655 individuals, selected by the partners of the MADO project. For comparison, expected microhaplotype frequencies based on the HLA-DRB1 frequencies provided by www.allelefrequencies.net are given in brackets.

The determined microhaplotype frequencies mirror the trend of the expected frequencies based on the allele frequencies of the HLA-DRB1 alleles in the Western European populations very well. Deviations from the expected microhaplotype frequencies are probably due to different sample size and different sampling strategy (different population).
198 samples out of these 655 were also HLA typed with SBT by two specialized HLA typing laboratories at the University Hospital in Vienna and the Centre Hospitalier Universitaire de Montpellier. For all the samples the estimated HLA types include the HLA type identified by SBT. These results are shown in the appendix (see section 5.5). In case of a green highlighted sample the HLA type with the highest likelihood matched the SBT identified HLA type. For yellow highlighted samples the HLA types found by SBT is in the list of estimated HLA types, but not the most likely. For red labelled samples no HLA type could be estimated. No discordant results were found.

3.5.1 HLA type estimation with missing data points

The maximally possible resolution of estimated HLA types for all samples is reached when all markers of a system yield results. In this case of HLA-DRB1 only results from maximally eight markers were included for the HLA type estimation. Since the markers HLADR_122r2 and DRB1_299f1 failed for all sample DNAs, the HLA type estimation was made with the results of the other eight markers. Further the markers DRB1_125r1 failed in 44 cases, DRB1_196f1 in 50 cases, DRB1_197r1 in 78 cases, DRB1_227f1 in 39 cases DRB1_261r1 in 162 cases, DRB1_308r1 in 56 cases and DRB1_341f1 in 14 cases. Missing results are mostly the result of failure of the extension reaction, the alkylation reaction and/or sample preparation on the MALDI-TOF sample carrier. Failure of the extension reaction was mainly observed for samples at the periphery of the microtitre plate. To avoid losing results from these samples, the assay was routinely carried out with two different plate arrangements, as outline in details in section 2.13 “Sample organisation for high throughput genotyping”.
Nevertheless for 95 % of the individuals an estimation of the HLA-DRB1 type could be made. For some individuals microhaplotyping results from just four markers sufficed to carry out the HLA type estimation. This is the case for the

individuals A00CG6O and A00CNK8 (see appendix section 5.5). These four markers were sufficient to identify one parental allele with a likelihood of >99 %.