Genes and Immunity (2006) 7, 487–493

www.nature.com/gene

ORIGINAL ARTICLE

Extensive multiallelic analysis of the relationship between HLA-DRB1 and rheumatoid arthritis using a Bayesian partition model

H Marotte[1,3], M Tournoud[2,3], M-A Cazalis[1], B Mougin[1], P Roy[2] and P Miossec[1]
[1]*Department of Immunology and Rheumatology, Unité Mixte Hospices Civils de Lyon – BioMérieux, Hôpital Edouard Herriot, Lyon Cedex, France and* [2]*Service de Biostatistiques HCL, Laboratoire de Biostatistique-Santé, Université Claude Bernard Lyon 1, UMR CNRS, Lyon Cedex, France*

To analyse the association between individual HLA-DRB1 locus genotypes and rheumatoid arthritis (RA) susceptibility, taking in account the multiallelic nature of the shared epitope (SE). In total, 538 patients and 536 controls were genotyped for 12 alleles of the HLA-DRB1 locus. A Bayesian partition model and multivariate logistic models were used to assess the role of the SE and of its individual components. The SE was associated with RA susceptibility (odds ratio (OR) 2 versus 0 SE copy = 9.99 (95 CI 4.69–15.30) and OR 1 versus 0 SE copy = 3.16 (95% CI 2.42–4.12)). The Bayesian partition model supplied a permutation of the HLA-DRBA locus alleles ordered by increasing disease risk. Alleles associated with highest risks are those that code for the SE. The individual OR estimations for the HLA-DRB1 locus genotypes went from OR = 1.00 (95% CI 1.00–1.25) for the less associated genotype to OR = 21.40 (95% CI 8.02–65.79) for the most associated one. In conclusion, the allele order risk and the OR estimations for individual genotypes of the HLA-DRB1 locus were consistent with the SE theory. Using an exploratory statistical method without a priori hypothesis, our study allowed a detailed analysis of the multiallelic nature of the SE.
Genes and Immunity (2006) **7,** 487–493. doi:10.1038/sj.gene.6364319; published online 6 July 2006

Keywords: *shared epitope; rheumatoid arthritis; multiallelic analysis*

Rheumatoid arthritis (RA) is a chronic inflammatory disease that affects 0.5–1% of the adult population worldwide. Its prevalence among women is approximately four times its prevalence among men.[1,2] However, RA is not equally aggressive in all patients, nor has the same outcome and response to treatment. Furthermore, many reports support the hypothesis of a genetic predisposition. The most documented issue is the implication of the highly polymorphic HLA-DRB1 locus that can also be regarded as a biallelic locus through the so-called 'shared epitope' (SE) theory.[3–7]

While the association between the SE and RA susceptibility is well established, the gene–dose effect is still controversial.[5,6,8,9] Furthermore, the effects of SE statistical interactions with age and sex on RA susceptibility are poorly documented. These differences could be due to lack of power or to 'population stratification', that is, different genotype frequencies between groups of a given population due to unique characteristics of these groups, such as common genetic and social histories, mating preferences, migration patterns, etc.[10] Inadequate statistical methods such as multiple subgroup analyses could also explain the result heterogeneity where a global analysis to test the gene–dose effect and interactions would be more adequate.

Few studies have considered the HLA-DRB1 locus as multiallelic when analyzing the association of different alleles or genotypes with RA susceptibility. Most studies focused on the alleles individually instead of considering the genotypic level, which is the correct measure of individual genetic exposure. Furthermore, comparing each allele versus all the others might lead to multiple testing problems, including the complex use of corrections such as the Bonferroni method.

In this report, a case–control design was applied to analyse the association between the SE and RA susceptibility. Several models of risk combinations were built to analyse the gene–dose effect and the gene–environment interactions. Furthermore, the association between the HLA-DRB1 locus genotypes and RA susceptibility was examined *via* a multiallelic approach. A Bayesian partition model[11] was used to assess the odds ratio (OR) of the different genotypes and to verify the *a priori* biological hypothesis of the SE.

Materials and methods

Patients and healthy controls

In total, 538 patients (cases) who met the criteria of the American College of Rheumatology 1987 for RA diag-

Correspondence: Professor P Miossec, Clinical Immunology Unit, Hopital Edouard Herriot, 69437 Lyon Cedex 03, France.
E-mail: miossec@univ-lyon1.fr
[3]These authors contributed equally to this work.
Received 29 March 2006; revised 17 May 2006; accepted 24 May 2006; published online 6 July 2006

nosis[12] were enrolled in the study (Table 1). All patients were resident in the Rhône-Alpes area, France. A cohort of 536 healthy volunteers (Table 1) from the same area was used as a control group. These volunteers, without chronic disease, were selected during their annual check-up at a work-related health organization (Centre ISBA, Lyon, France). Sex, age, and genetic data were collected for patients and controls. The protocol of the study was approved by the local ethics committee and all subjects provided a written informed consent for the genetic analysis.

Table 1 Characteristics of the study RA patients and healthy controls

Characteristics	*Controls*	*RA patients*
Number of subjects (Females %)	536 (24.25)	538 (75.47)
Mean age in years at inclusion for controls and disease onset for patients	48.4 [46.0–52.0][a]	42.28 [32.0–52.0]
Mean disease duration in months		120 [36.0–180.0] (473)[b]
SE genotypes		
No copy of the SE %	64.06	37.77
One copy of the SE %	31.71	44.81
Two copies of the SE %	4.23	17.42

Abbreviations: RA, rheumatoid arthritis; SE, shared epitope.
[a][1st quartile–3rd quartile].
[b]() indicates the number of controls or patients for which the information was available.

Polymorphism gene typing

Genomic DNA was extracted from 200-μl of peripheral whole blood from patients and controls using the QIAamp DNA Blood Mini Kit (Qiagen GmbH, Hilden, Germany). Genotyping was performed by Enzyme-Linked OligoSorbent Assay (bioMérieux Assay ELOSA) as previously described.[13] Specific features of the test are reported below.

Exon 2 regions of both HLA-DR and HLA-B were polymerase chain reaction (PCR) amplified using a combination of DR-specific and B-specific primers. The DR-specific primer sequences were: Forward primers: 5′-GTT CTT GTC CCC CCA GCA CG-3′ and 5′-GTT CTT GTC CCC ACA GCA CG-3′; Reverse primer: 5′-TCG CCG CTG CAC TGT GAA G-3′. The B-specific primer sequences were: Forward primer: 5′-GGG AGG AGC GAG GGG ACC GCA-3′; Reverse primer: 5′-ATC TCG GAC CCG GAG ACT CG-3′. To further classify some samples, DR1-specific primers were used. The DR1-specific primer sequences were: Forward primers: 5′-GGC AGC TTA AGT TTG AAT G-3′; Reverse primer: 5′-TCG CCG CTG CAC TGT GAA G-3′. The amplification mixture was composed of: 50 mM Tris-HCl (pH 8.8), 15 mM ammonium sulphate, 1.5 mM $MgCl_2$, 50 μM EDTA, 0.01% (w/v) gelatine, 0.2 mM dNTPs, 2.5 U AmpliTaq (Perkin-Elmer, Boston, MA, USA), 0.15 μM for HLA-DR primers, 0.3 μM for HLA-DR4 primer, and 0.4 μM for HLA-B primers in a 100 μl volume reaction. Fifty to

Table 2 Relationship between ELOSA capture probes and genotype notations

Amino-acid sequence	*ELOSA capture probes*		*Genotype notation: HLA-DRB1*
QKRAA[a]	0401	HLA DRB1*	04011, 04012, 0413, 0416, 0421, 0426, 0433, 0434, 0435, 0438, 0409, 1419, 1421
QRRAA[a]	01S	HLA DRB1*	0101, 01021, 01022, 0103, 0104, 0105, 0106, 0107, 0108, 0103
QRRAA[a]	0404	HLA DRB1*	0404, 0408, 0419, 0423, 0440, 0442
QRRAA[a]	0405	HLA DRB1*	04051, 04052, 0410, 0428, 0429, 0430
RRRAA[a]	10	HLA DRB1*	10011, 10012
DERAA[b]	0402	HLA DRB1*	0402, 0414, 0437, 0103, 1510, 1102, 1103, 1111, 1114, 1116, 1120, 1121, 1136, 1140, 1141, 13011, 13012, 13021, 13022, 1308, 1315, 1316, 1319, 1320, 1322, 1323, 1324, 1327, 1328, 1329, 1331, 1334, 1335, 1336, 1339, 1340, 1341, 1343, 1345, 1304, 1332, 1338, 1348, 1317, 1416
QKRAE[b]	0403	HLA DRB1*	04031, 04032, 0406, 04071, 04072, 0420, 0427, 0439, 0441, 0411, 0417, 1433
Various amino-acid sequences[b]	2+7+9	HLA DRB1*	02 (15 and 16), 1510, 07011, 07012, 0703, 0705 and 09012
Various amino-acid sequences[b]	03	HLA DRB1*	03 (except 0312, 03022, 0314, 0315, 0317)
Various amino-acid sequences[b]	52	HLA DRB1*	03 (except 0312, 03022, 0314, 0315, 0317), 03022, 0314, 0315, 0317, 0312, 0820, 11011, 11012, 11013, 11041, 11042, 1106, 1107, 11081, 11082, 1109, 1110, 1112, 1113, 1115, 1117, 1118, 1119, 1123, 1124, 1125, 1127, 1128, 1129, 1131, 1132, 1133, 1135, 1136, 1142, 1102, 1103, 1111, 1114, 1116, 1120, 1121, 1136, 1140, 1141, 1105, 13011, 13012, 13021, 13022, 1308, 1315, 1316, 1319, 1320, 1322, 1323, 1324, 1327, 1328, 1329, 1331, 1334, 1335, 1336, 1339, 1340, 1341, 1343, 1345, 13031, 13032, 1312, 1313, 1321, 1330, 1333, 1349, 1304, 1332, 1338, 1348, 1305, 1306, 13071, 13072, 1309, 1310, 1311, 1314, 1318, 1325, 1326, 1337, 1342, 1344, 1346, 1347
Various amino-acid sequences[b]	8+12	HLA DRB1*	0801, 08032, 0805, 0806, 0810, 0812, 0816, 0817, 0818, 0822, 0823, 08021, 08022, 08041, 08042, 08043, 0807, 0808, 0809, 0811, 0813, 0815, 0819, 0824, 1105, 12011, 12012, 12021, 12032, 1204, 1205, 1206, 1207, 1208, 1317, 1404, 1411, 1415, 1428, 1431
Various amino-acid sequences[b]	01N	HLA DRB1*	01022, 0106, 0103

Abbreviation: SE, shared epitope.
[a]Coding for the SE.
[b]Not coding for the SE.
2+7+9 means a probe specificity for the HLA DRB1*01, DRB1*07 and DRB1*09.
8+12 means a probe specificity for the HLA DRB1*08 and DRB1*12.

200 ng of extracted DNA were used per amplification. Cycling conditions were as follows: 2 min denaturation at 95°C, then 4 cycles with 30 s at 95°C, 30 s at 68°C, 30 s at 72°C, then 4 cycles with 30 s at 95°C, 30 s at 57°C and 30 s at 72°C, then 3 cycles with 30 s at 95°C, 30 s at 64°C and 30 s at 72°C, then 30 cycles with 30 s at 95°C, 30 s at 60°C and 30 s at 72°C, then 7 min at 72°C. PCR efficiency was checked by agarose gel electrophoresis. Amplicons were hybridized on specific capture probes coated in eight-well strips assembled on a microtitre plate frame, followed by semiautomated washing, colorimetric detection, and reading.

Genetic data

The genetic data available corresponded to the genotyping results of the highly polymorphic locus for HLA-DRB1, identifying 12 different groups of alleles. The corresponding genotypes detected by the ELOSA probes are detailed in Table 2. The HLA typing assay for

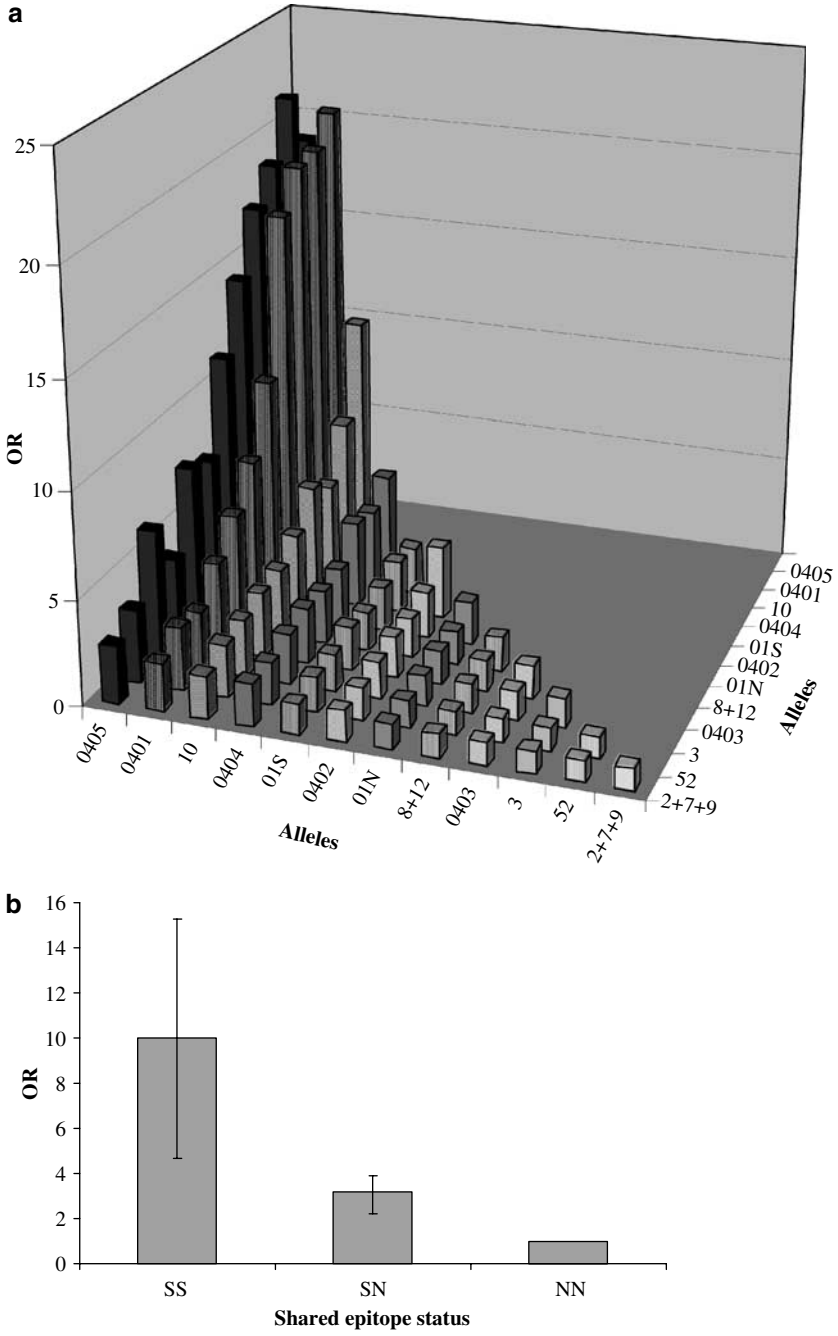

Figure 1 OR for RA susceptibility according to the different genotypes (**a**) or SE status (**b**). (**a**) OR estimations for the association between RA susceptibility and individual genotypes, using 2 + 7 + 9/52 as reference. (**b**) OR estimations and 95% confidence intervals for the association between RA susceptibility and the SE status, using 0 copy of the SE as reference.

the SE uses generic detection probes and specific genotyping probes (Table 2). Three genotypes were observed according to the number of copies of the SE: no copy, one copy, and two copies. The HLA typing assay was unable to differentiate homozygote patients for 01S and heterozygote patients for 01S and 01N (HLA-DRB1*0106). In Caucasian populations, the HLA-DRB1*0106 was only found extremely rarely in Argentina or Azores populations (http://www.allelefrequencies.net).[14] Thus, it was considered that ambiguous cases and controls were homozygote for 01S/01S.

Statistical analysis

Study of the SE. To analyse the SE association with RA susceptibility, SE genotypes were compared between patients and controls. Owing to an imbalance regarding sex and age in the control population, a systematic adjustment was performed on these confounders. SE × age and SE × sex statistical interactions were tested at the 0.10 significance level.

Five multivariate logistic models were built to analyse the association between the SE genotypes and the disease risk providing estimations of OR in comparison with no copy for the SE.[15] The recessive model implies that the disease risk for one copy of the SE is equal to the disease risk for no copy of the SE. The dominant model implies that the disease risk for two copies of the SE is equal to the disease risk for one copy of the SE and the multiplicative model implies that the OR for two copies of the SE is the square of the OR for one copy of the SE. Likelihood ratio tests (LR Test) were performed to compare nested models.[16] Comparisons between recessive, dominant, multiplicative, and the null model (identical disease risk for the three SE statuses) were used to test the association between the SE and RA susceptibility. Comparisons with the full model (equally estimated disease risks and observed disease risks) were used to test the gene–dose effect. A *P*-value smaller than 0.05 was considered as significant.

Study of the HLA-DRB1 locus

Allelic association study: T3 and T4 Sham tests. Considering the HLA-DRB1 locus as a multiallelic locus, equality of allelic frequencies was tested between patients and controls with T3 and T4 Sham tests.[17] These tests were used to avoid problems of multiple testing and cells with small numbers in large contingency tables. T3 and T4 tests were based on classification of alleles into two groups. The T3 test compared between patients and controls the distribution of each allele versus all other alleles, whereas the T4 test compared the distribution of all possible combinations of alleles versus all others. T3 and T4 statistics tested the same null hypothesis; that is, a global homogeneity of allelic frequencies between patients and controls. However, under statistical significance and the hypothesis of two groups of alleles, the T4 statistic provided a binary partition of the alleles with one group being more associated with RA susceptibility than the other group. This partition was informally compared to the SE allelic combination.

Genotypic association study: seaman's Bayesian partition model. A Bayesian partition model[11] was used to study the association between RA susceptibility and the individual genotypes of HLA-DRB1 locus. This model is particularly efficient for the analysis of highly polymorphic disease susceptibility genes. As the number of subjects having each genotype was low, analyzing each genotype separately was limited by small genotype sample sizes and multiple testing problems. The Bayesian partition model is well adapted to study small genotype sample sizes because it clusters the genotypes according to risk. Clustering genotypes enables to circumvent the size problem because genotypes with small numbers of subjects are grouped with others. Although the Bayesian partition model clusters the genotypes, it supplies one disease risk estimation (posterior OR estimation) per genotype. Bayesian methods lead to *a posteriori* (posterior) probability computations that combine the *a priori* information (knowledge that practitioners have before the analysis) with the current information in the data (observations). In the present case, little information on each genotype associated disease risk was available, thus no *a priori* information was added. The Bayesian partition model also supplies an allelic order risk that corresponds to a

Table 3 HLA-DRB1 locus genotypes (A1/A2). Distribution for patients and controls (numbers in parenthesis)

A2	*A1*											
	0405	*0401*	*10*	*0404*	*01S**	*0402*	*01N**	*8+12*	*0403*	*03*	*52*	*2+7+9*
0405	0 (1)											
0401	11 (0)	10 (1)										
10	2 (0)	3 (0)	0 (0)									
0404	2 (0)	16 (1)	1 (1)	1 (2)								
01S*	6 (0)	18 (0)	3 (1)	9 (2)	7 (11)							
0402	2 (0)	2 (0)	0 (0)	0 (0)	2 (2)	0 (0)						
01N*	0 (0)	4 (0)	1 (0)	0 (1)	0 (0)	0 (0)	0 (0)					
8+12	3 (0)	7 (1)	1 (0)	2 (1)	2 (6)	0 (1)	0 (1)	4 (2)				
0403	2 (2)	3 (1)	0 (0)	1 (0)	5 (1)	0 (0)	0 (0)	1 (1)	1 (1)			
03	4 (0)	7 (9)	3 (1)	5 (1)	8 (9)	0 (1)	0 (2)	5 (7)	2 (1)	14 (10)		
52	4 (1)	29 (11)	6 (1)	8 (6)	34 (29)	2 (4)	1 (3)	9 (17)	2 (6)	14 (26)	32 (49)	
2+7+9	12 (6)	19 (17)	6 (5)	19 (8)	28 (31)	4 (2)	2 (4)	9 (14)	2 (4)	22 (28)	44 (80)	23 (39)

01N*: 01022, 0106, 0103.

01S*: 0101, 01021, 0104, 0105, 0107, 0108, 0103.

The numbers of controls per each genotype are indicated in brackets.

Table 4 HLA-DRB1 locus genotypic (A1/A2). Posterior OR estimation (95% credibility intervals) for RA susceptibility

A2	*A1*												
	0405	*0401*	*10*	*0404*	*01S**	*0402*	*01N**	*8+12*	*0403*	*03*	*52*	*2+7+9*
0405	18.79 (3.65–63.30)											
0401	21.40 (8.02–65.79)	20.89 (7.61–64.58)										
10	18.70 (3.82–62.10)	19.63 (6.09–63.24)	11.31 (1.96–53.56)									
0404	17.22 (3.31–60.03)	19.45 (7.10–60.00)	7.08 (2.02–33.69)	4.70 (1.77–21.51)								
01S*	14.45 (2.80–57.21)	17.74 (5.16–58.62)	4.74 (1.88–25.19)	3.74 (1.77–14.52)	2.25 (1.17–3.91)							
0402	11.33 (1.96–55.65)	10.45 (1.83–53.51)	5.55 (1.28–39.36)	4.08 (1.14–29.12)	2.46 (1.00–6.06)	3.49 (1.00–38.03)						
01N*	7.12 (1.66–47.50)	7.36 (1.70–47.38)	4.05 (1.06–32.96)	2.69 (1.00–12.85)	2.12 (1.00–6.15)	2.19 (1.00–17.99)	2.06 (0.68–26.51)					
8+12	7.61 (1.86–46.86)	5.61 (1.93–33.26)	3.26 (1.23–19.46)	2.53 (1.03–9.23)	1.83 (1.00–3.23)	1.88 (1.00–4.17)	1.58 (1.00–3.44)	1.62 (1.00–3.53)				
0403	4.06 (1.60–20.60)	4.16 (1.51–23.04)	3.07 (1.07–17.84)	2.65 (1.00–13.18)	2.08 (1.00–5.08)	1.94 (1.00–6.29)	1.60 (1.00–4.56)	1.50 (1.00–3.37)	1.58 (0.99–9.33)			
03	6.34 (1.63–45.25)	2.75 (1.42–5.37)	2.71 (1.13–12.85)	2.37 (1.07–6.78)	1.76 (1.00–3.19)	1.73 (1.00–3.68)	1.43 (1.00–3.08)	1.39 (1.00–2.63)	1.40 (1.00–2.97)	1.44 (1.00–2.80)		
52	3.50 (1.65–16.87)	2.99 (1.79–5.88)	2.52 (1.29–10.19)	2.01 (1.14–3.81)	1.68 (1.00–2.66)	1.52 (1.00–2.91)	1.28 (1.00–2.50)	1.16 (1.00–1.95)	1.15 (1.00–2.12)	1.08 (1.00–1.65)	1.04 (1.00–1.44)	
2+7+9	2.75 (1.53–5.93)	2.29 (1.32–3.89)	2.03 (1.00–3.93)	2.03 (1.16–3.73)	1.43 (1.00–2.47)	1.49 (1.00–2.97)	1.21 (1.00–2.41)	1.13 (1.00–1.95)	1.11 (1.00–2.01)	1.09 (1.00–1.82)	1.00 (1.00–1.00)	1.01 (1.00–1.25)

Abbreviations: OR, odd ratios; SE, shared epitope.
The 2+7+9/52 genotype was taken as a reference for OR estimations, integrating all distribution numbers in the controls and patients as shown in Table 3.
01N*: 01022, 0106, 0103.
01S*: 0101, 01021, 0104, 0105, 0107, 0108, 0103.

classification of the alleles by increasing disease risk. The allelic order risk is more informative than a binary partition of the alleles, like the SE hypothesis. However, the binary partition obtained by determining an appropriate cutting point will be compared to the SE hypothesis. A more detailed description of the model is supplied in Annex A, for interested readers.

Results

The SE and RA susceptibility

The genotype distribution of the included RA patients and healthy controls is summarized in Table 1. According to the multivariate analysis adjusted on age and sex, the SE was associated with RA susceptibility. The SE was highly associated with RA risk, whatever the statistical model used. The difference between the full and the multiplicative model was not significant (LR Test = 0.65, $P = 0.55$). Accordingly, the multiplicative model was selected. By construction, this model implied a high gene–dose effect. Taking no copy of the SE as a reference, the disease risk adjusted for age and sex appeared 9.99-fold higher for two copies of the SE (OR = 9.99 (95% CI 4.69–15.30)) and 3.16-fold higher for one copy of the SE (OR = 3.16 (95% CI 2.42–4.12)) (Figure 1). There were interactions neither between the SE and age ($P = 0.98$) nor between the SE and sex ($P = 0.36$).

The HLA-DRB1 locus and RA susceptibility

Both T3 and T4 Sham tests rejected the null hypothesis of identical allelic frequencies for patients and controls ($P < 10^{-3}$ in both cases). The latter results showed that the HLA-DRB1 locus was associated with RA susceptibility because the repartition of the HLA-DRB1 locus alleles was not the same for patients and controls. Considering the hypothesis of a binary subtyping (i.e., the existence of two groups of alleles: susceptibility and neutrality ones), the statistic T4 provided useful information to classify the alleles into two groups. It classified DRB1*0401, 0404, 0405, 10 as a susceptibility allele group. These results were consistent with the SE *a priori* hypothesis except for the 01S allele that was noted as a susceptibility allele in the latter classification.

Seaman's Bayesian partition model was used to study the association between individual genotypes of the HLA-DRB1 locus and RA susceptibility. The distribution of the HLA-DRB1 locus genotypes for patients and controls is given in Table 3. Using the 2 + 7 + 9/52 genotype as a reference (Table 2), after running 100 000 iterations, Seaman's Bayesian partition model provided an OR estimation per genotype with its corresponding 95% credibility interval (Tables 3 and 4 and Figure 1). The individual OR estimations for the HLA-DRB1 locus genotypes went from OR 2 + 7 + 9/2 + 7 + 9 versus 2 + 7 + 9/52 = 1.00 (95% CI 1.00–1.25) to OR 0405/0401 versus 2 + 7 + 9/52 = 21.40 (95% CI 8.02–65.79).

The allele order risk that led to the partition model was also very informative. Table 5 presents the posterior probabilities of each allele of being in each position of the allele order risk and mean ranks. The allele order risk, built with mean ranks, corresponds to a classification of the alleles of the HLA-DRB1 locus by increasing disease risk. The allele with the minimum posterior mean position (2 + 7 + 9) may be associated with the lowest

Table 5 Posterior probabilities for each allele of being at each rank in the risk order and mean ranks

Rank	*Posterior probabilities of allele order*											
	2+7+9	*52*	*03*	*0403*	*8+12*	*01N*	*0402*	*01S*	*0404*	*10*	*0401*	*0405*
1	0.39	0.16	0.10	0.10	0.08	0.13	0.04	0.00	0.00	0.00	0.00	0.00
2	0.33	0.29	0.12	0.09	0.06	0.06	0.03	0.00	0.00	0.00	0.00	0.00
3	0.19	0.31	0.15	0.16	0.08	0.07	0.04	0.00	0.00	0.00	0.00	0.00
4	0.07	0.18	0.24	0.19	0.15	0.10	0.05	0.01	0.00	0.00	0.00	0.00
5	0.01	0.05	0.23	0.17	0.26	0.13	0.09	0.05	0.00	0.00	0.00	0.00
6	0.00	0.01	0.13	0.11	0.24	0.17	0.13	0.18	0.02	0.01	0.00	0.00
7	0.00	0.00	0.02	0.07	0.10	0.13	0.17	0.40	0.08	0.04	0.00	0.00
8	0.00	0.00	0.00	0.07	0.03	0.10	0.17	0.30	0.24	0.09	0.00	0.01
9	0.00	0.00	0.00	0.02	0.01	0.07	0.16	0.06	0.41	0.21	0.02	0.03
10	0.00	0.00	0.00	0.01	0.00	0.04	0.11	0.00	0.18	0.45	0.12	0.09
11	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.05	0.13	0.56	0.25
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.07	0.29	0.62
Mean rank	1.98	2.69	3.87	4.39	4.72	5.28	6.85	7.13	8.83	9.69	11.12	11.43

risk of disease, whereas the allele with the maximum posterior mean position (0405) may be associated with the greatest risk of disease. Furthermore, the allele order risk was split into two groups to verify the SE hypothesis. The exact SE classification ((2 + 7 + 9, 52, 03, 0403, 8 + 12, 01N, 0402) versus (01S, 0404, 10, 0401, 0405)) was obtained by cutting the permutation between 0402 and 01S.

Discussion

This study used a case–control design to assess the effects of the SE genotypes on RA susceptibility. We have reliably applied some classical models of epidemiology to a genetic association study and used different models of risk combinations (dominant, recessive, multiplicative) to analyse a possible gene–dose effect. These multivariate models could take into account variables, such as age and sex, and test possible interactions with sex and age at disease onset.

Our results strengthened the previously described association between RA susceptibility and the SE ((5), (6), (8), (9)) for a Caucasian population of the Rhône-Alpes area. The multiplicative model was selected, indicating a clear gene–dose effect. By construction, the OR for two copies of the SE (OR = 9.99 (95% CI 4.69–15.30)) was the square of the OR for one copy of the SE.

Although the role of the SE in RA susceptibility remained controversial (gene–dose effect, interactions with age and sex), the main purpose of this study was not to analyse with multivariate statistical models the global effect of the SE. As a result of the multiallelic structure of the HLA-DRB1 locus, it was more appealing to refine the contribution of the SE components. Consequently, we have examined the association between RA susceptibility and all the HLA-DRB1 locus genotypes. The Bayesian partition model provided an OR estimation for each genotype of the HLA-DRB1 locus. Furthermore, the model supplied an interesting allele order risk: (2 + 7 + 9; 52; 8 + 12; 03; 0403; 01N; 0402; 01S; 0404; 10; 0401; 0405). Under the hypothesis of two groups of alleles associated with RA susceptibility, the *a priori* SE theory (4) which was raised on a biological hypothesis and on molecular considerations was verified. To our knowledge, this is the first time that an OR estimation per genotype (with its 95% credibility interval) is given. For a locus with 12 alleles, there are 66 possible genotypes. Hence, it was not possible to compare directly all the genotypes between patients and controls and obtain an accurate OR estimation. Indeed, comparing each genotype versus all the others to assess a genotype risk was difficult because of the genotype sample sizes and because this could lead to nonindependent multiple tests. By performing a global analysis, the Bayesian partition model is a good approach to analyse polymorphic disease genes. In addition, to estimate the OR associated to a particular genotype, the model used information of all the genotypes sharing one allele with the considered genotype. As an example, if we consider the genotype 0405/0405, no patient and one only controls were homozygote for 0405; the estimated OR was 18.79 (95% CI 3.65–63.30). As the 0405/0405 homozygote genotype was very rare, the model used the fact that 0405 was the allele associated with the highest risk to compute an OR estimation for 0405/0405 homozygotes. The underlying assumption about the combined effect of the two alleles making up a genotype was genetically plausible and permitted a highly flexible model.

In conclusion, using updated statistical analyses, we have clearly shown that the SE is a susceptibility marker for RA in a Caucasian population. Looking into the SE components, Seaman's Bayesian partition model provided disease risk estimation with its 95% credibility interval for each genotype of the HLA-DRB1 locus. Furthermore, using an exploratory statistical method without an *a priori* hypothesis and taking into account its multiallelic nature, we have extended the SE biological theory and that a SE-positive status remains a strong but complex genetic indicator for RA.

Acknowledgements

We thank Shaun Seaman who kindly gave us the source code (C language) of the Bayesian partition model.

References

1 Firestein GS. Evolving concepts of rheumatoid arthritis. *Nature* 2003; **423**: 356–361.
2 Gabriel SE. The epidemiology of rheumatoid arthritis. *Rheum Dis Clin North Am* 2001; **27**: 269–281.
3 Harney S, Wordsworth BP. Genetic epidemiology of rheumatoid arthritis. *Tissue Antigens* 2002; **60**: 465–473.
4 Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis. An approach to understand the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987; **30**: 1205–1213.
5 De Vries N, Tijssen H, van Riel PL, van de Putte LB. Reshaping the shared epitope hypothesis: HLA-associated risk for rheumatoid arthritis is encoded by amino acid substitutions at positions 67–74 of the HLA-DRB1 molecule. *Arthritis Rheum* 2002; **46**: 921–928.
6 Gonzalez-Gay MA, Garcia-Porrua C, Hajeer AH. Influence of human leukocyte antigen-DRB1 on the susceptibility and severity of rheumatoid arthritis. *Semin Arthritis Rheum* 2002; **31**: 355–360.
7 Gorman JD, Lum RF, Chen JJ, Suarez-Almazor ME, Thomson G, Criswell LA. Impact of shared epitope genotype and ethnicity on erosive disease. meta-analysis of 3240 rheumatoid arthritis patients. *Arthritis Rheum* 2004; **50**: 400–412, 2004.
8 Toussirot E, Auge B, Tiberghien P, Chabod J, Cedoz JP, Wendling D. HLA-DRB1 alleles and shared amino acid sequences in disease susceptibility and severity in patients from eastern France with rheumatoid arthritis. *J Rheumatol* 1999; **26**: 1446–1451.
9 Thomson W, Harrison B, Ollier B, Wiles N, Payton T, Barrett J *et al*. Quantifying the exact role of HLA-DRB1 alleles in susceptibility to inflammatory polyarthritis: results from a large, population-based study. *Arthritis Rheum* 1999; **42**: 757–762.
10 Huizinga TWJ, Pisetsky DS, Kimberly RP. Associations, populations, and the truth. Recommendations for genetic associations studies in Arthritis & rheumatism. *Arthritis Rheum* 2004; **50**: 2066–2071.
11 Seaman SR, Richardson S, Stücker I, Benhamou S. A bayesian partition model for case-control studies on highly polymorphic candidate genes. *Genet Epidemiol* 2002; **22**: 356–368.
12 Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988; **31**: 315–324.
13 Mougin B, Garnero P, Borel O, Compagnon C, Barbalat V, Marotte H *et al*. A routine assay for the direct analysis of HLA-DR-related shared epitope and B27 alleles in chronic inflammatory arthritis. *J Immunol Methods* 2001; **256**: 47–53.
14 Collins MM, Tang T, Slack R, Sintasath D, Hartzman RJ, Ng J *et al*. The relative frequencies of HLA-DRB1*01 alleles in the major US populations. *Tissue Antigens* 2000; **55**: 48–52.
15 McCullagh P, Nelder JA. *Generalized Linear Models*, 2nd edn. Chapman and Hall: London, 1989.
16 Breslow NE, Day NE. *Statistical Methods in Cancer Research Vol. 1 – The Analysis of Case-Control Studies* ed. IARC: Lyon, 1980.
17 Sham PC, Curtis D. Monte Carlo tests for association between disease and alleles at highly polymorphic loci. *Ann Hum Genet* 1995; **59**: 97–105.

Appendix A

Details on Seaman's Bayesian partition model

The basic assumptions of the partition model are that all possible genotypes of the HLA-DRB1 locus belong to a number of groups, that all genotypes in the same group convey the same risk of disease for the positive individuals and that genotypes in different groups convey different risks. A genotype partition is built from an allele order risk that corresponds to a classification of the alleles by increasing disease risk. Each genotype partition satisfies a particular assumption: if two genotypes share one allele but differ for the other, then the risk conveyed by a genotype with a higher allele in the order risk cannot be less than the risk conveyed by a genotype with a lower allele in that order risk. Accordingly, if an allele A_j is higher than the allele A_i in the allele order risk, whatever the allele A_k, the risk associated with the A_j/A_k genotype is greater than the risk associated with A_i/A_k. Although the Bayesian partition model clusters the genotypes, it supplies one disease risk estimation per genotype. Indeed, at each iteration of the fitting algorithm, an allele order risk, a partition model and group disease risks are sampled from their posterior distribution. The posterior distribution of each genotype risk is then recovered from the group disease risk the genotype belonged to at each iteration.

Uniform priors were placed on the allele order risk and the partition model. The prior distribution of the log (disease risk) of each group was the normal distribution $N(0, \sigma^2)$, $\sigma^2 = 2.34$. This value of σ^2 implied the assumption that 95% of groups of genotypes had disease risk in the range (0.05–20).

Results were based on 100 000 iterations. OR estimation and 95% credibility intervals were supplied for each genotype. 95% credibility intervals were constructed from the posterior distributions and could be interpreted as the classical 95% confidence intervals. The posterior probabilities of each allele to be at each rank as well as the mean rank of each allele were also computed. The mean rank of each allele provided the retained allele order risk. Results are presented in Table 5.