European Journal of Human Genetics (2003) 11, 794–801

www.nature.com/ejhg

ARTICLE

Analysis of the French National Registry of unrelated bone marrow donors, using surnames as a tool for improving geographical localisation of HLA haplotypes

Anna Degioanni*[,1,2], Pierre Darlu[2] and Colette Raffoux[3]

[1]*UMR 6578 CNRS/Université de la Méditerranée, Marseille, France;* [2]*INSERM U535, Villejuif, France;* [3]*France Greffe de Moelle, Hôpital Saint-Louis, Paris, France*

The first statistical analysis of the French National Registry of volunteer bone marrow donors estimated the probabilities of haplotype frequencies separately for each of the 20 administrative regions of France. Here we propose to use donors' surnames to increase the accuracy of location of the donor's geographical origin. This approach allows us to estimate haplotype frequencies for administrative entities (90 departments) smaller than regions and to correct for bias resulting from recent mobility. We analysed 30 777 donors typed for HLA-A,B and 17 745 donors typed for HLA-A,B,DR,DQ. By using the donors' surnames, we identified common and rare haplotypes (those found in only one department) and estimated the degree of HLA polymorphism at the department level. We also identified departments with a distinctive genetic structure (for example, Paris, Corsica, Pyrenees and Meurthe-et- Moselle). By providing a more accurate geographical distribution of HLA polymorphisms in France, this study will enable us to optimise policies for recruiting bone marrow donors and to improve the fit between the donor file and patients' needs.
European Journal of Human Genetics (2003) **11,** 794–801. doi:10.1038/sj.ejhg.5201038

Keywords: surname; HLA haplotype frequency; bone marrow donor registry

Introduction

In recent years, various national registries of volunteer bone marrow donors have been analysed increasingly often,[1–8] stimulated by the growing number of registered donors, the diversity of their geographical origins and the accuracy of the haplotype definitions in HLA genetic systems. These analyses are also required because of the increasing number of transplants and the need to improve the recruitment of volunteer donors, particularly by identifying the areas where specific donors are most likely to be found and where rare haplotypes occur more often than expected by chance.

In 1995, Lonjou *et al*[9] carried out the first study of the French National Registry of Volunteer Bone Marrow Donors. Their study estimated haplotype frequencies for the 20 administrative regions of France and thus enabled them to identify the regions where uncommon haplotypes have the highest probability of being found. The donors' geographic origins, however, were assigned very approximately, based as they were only on the area or town where the typing laboratory was located. The high mobility of French residents in recent decades means that this crude definition of origin can be highly erroneous. Most people were registered, at birth, in a town with a maternity ward or in a city where their parents moved to find work, and not necessarily where their relatives or other people likely

*Correspondence: Dr A Degioanni, Service d'Anthropologie Biologique, UMR 6578 CNRS/Université de la Méditerranée, faculté de Médecine, 27, bd Jean Moulin, 13385 Marseille cedex 05, France. Tel: +33 491324567; Fax: +33491324585; E-Mail: Anna.Degioanni@medecine.univ-mrs.fr
Received 1 May 2002; revised 4 April 2003; accepted 16 April 2003

to share haplotypes with them are settled. The best way to find the donors' true origin would be to determine the genealogical history of each donor, to maximise the chances of finding shared genetic structures, but the size of the registry (currently more than 100 000 volunteer donors) makes this option unfeasible.

As an alternative, we propose the use of donor surnames to increase the accuracy of the inference of donors' geographical origins. The basic idea is that (i) surnames are transmitted through the father's line, together with the genes of the Y chromosome, (ii) most (but not all) surnames have a fairly precise geographic origin, and (iii) the origins of male and female surnames are highly correlated, since the marriage circle in recent decades in France has not exceeded 5 km on average.[10] In this study, we used the surnames of male donors to estimate haplotype frequencies for the 90 French departments (Figure 1 reports the department names and numbers), instead of the 20 regions (as in Lonjou *et al*[9]).

Materials and methods

Donors, HLA and surnames

The French file of volunteer bone marrow donors contains 40 104 men, all serologically defined for HLA-A and HLA-B. Among them, 25 455 are also defined for both HLA-DRB1 and HLA-DQB1, and 22 211 only for HLA-DRB1.

We knew the name and location of the tissue-typing laboratory for each donor as well as his surname. Women were excluded from the analysis because it was impossible to be certain if they were registered under their father's or their husband's surname.

Information about French surname distribution is based on data obtained from the 'Births Registry' provided by Institut National des Statistiques et Etudes Economiques (INSEE), the French agency that collects, analyses, and disseminates economic and statistical information. This register contains, for each of the 36 000 communes in France and for every surname, the number of infants born during two separate periods (1891–1915, P1, and 1916–1940, P2) and alive in 1970. Data from towns and smaller municipalities in the same department are pooled together, and surname frequencies calculated for each of the 90 French departments.

Several investigations ultimately revealed that some donor surnames were absent from the INSEE registry for the first period (P1). These names represent immigrants into France since the first World War, from more than 40 different countries. To minimise bias in estimating haplotype frequencies of the French population, they are excluded from the present analysis: further detailed analyses will be devoted to them. Of the donor surnames listed in the INSEE register, a few have been found in only one department and thus have an unambiguous geographical origin. Even though most surnames are present in

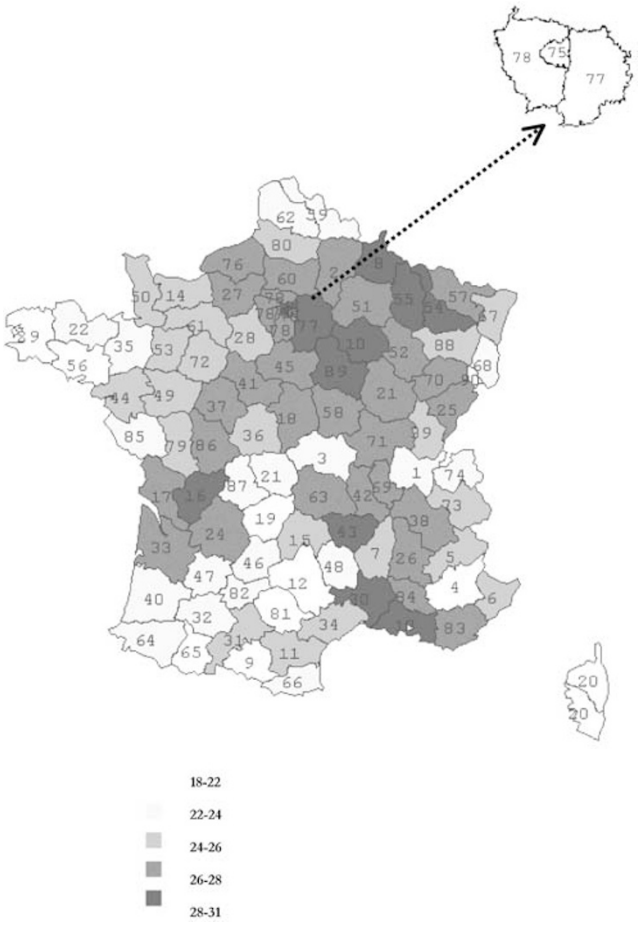

Figure 1 Departmental distribution of the minimum number of HLA-A, B haplotypes needed to reach 50% of the total frequency (Corsica, dept. no. 20, is dropped down from its real geographical position). Department number: 1 = Ain, 2 = Aisne, 3 = Allier, 4 = Alpes Hte Provence, 5 = Hautes Alpes, 6 = Alpes Maritimes, 7 = Ardèche, 8 = Ardennes, 9 = Ariège, 10 = Aube, 11 = Aude, 12 = Aveyron, 13 = Bouches du Rhône, 14 = Calvados, 15 = Cantal, 16 = Charente, 17 = Charente Maritime, 18 = Cher, 19 = Corrèze, 20 = Corsica, 21 = Côte d'Or, 22 = Côte d'Armor, 23 = Creuse, 24 = Dordogne, 25 = Doubs, 26 = Drôme, 27 = Eure, 28 = Eure et Loir, 29 = Finistère, 30 = Gard, 31 = Haute Garonne, 32 = Gers, 33 = Gironde, 34 = Hérault, 35 = Ille et Vilaine, 36 = Indre, 37 = Indre et Loire, 38 = Isère, 39 = Jura, 40 = Landes, 41 = Loir et Cher, 42 = Loire, 43 = Haute Loire, 44 = Loire atlantique, 45 = Loiret, 46 = Lot, 47 = Lot et Garonne, 48 = Lozère, 49 = Maine et Loire, 50 = Manche, 51 = Marne, Marne, 52 = Haute Marne, 53 = Mayenne, 54 = Meurthe et Moselle, 56 = Morbihan, 57 = Moselle, 58 = Nièvre, 59 = Nord, 60 = Oise, 61 = Orne, 62 = Pas de Calais, 63 = Puy Puy du Dôme, 64 = Pyrénées atlantiques, 65 = Hautes Pyrénées, 66 = Pyrénées orientales, 67 = Bas Rhin, 68 = Haut Rhin, 69 = Rhône, 70 = Haute Saône, 71 = Saône et Loire, 72 = Sarthe, 73 = Savoie, 74 = Haute Savoie, 75 = Paris, 76 = Seine maritime, 77 = Seine et Marne, 78 = Yvelines, 79 = Deux Sèvres, 80 = Somme, 81 = Tarn, 82 = Tarn et Garonne, 83 = Var, 84 = Vaucluse, 85 = Vendée, 86 = Vienne, 87 = Haute Vienne, 88 = Vosges, 89 = Yonne, 90 = Belfort.

more than one department, they may nevertheless be informative if the geographic centre of their maximum frequency is well localised and if this frequency decreased rapidly from there.

Most French surnames, however, are widespread in France (eg, Martin, Bernard, Durand, Dupont) or are present in several departments without a clear centre of origin[11] The present method (see surname method paragraph) allows us to overcome this difficulty, by providing us with the probability that a given surname comes from a given department, in view of its frequency in this department and in others.

The INSEE register included 17 971 donor surnames (names thus long established in France) that corresponded to a total of 30 777 donors. These donors corresponded to 8558 different HLA-A, B phenotypes and to 13 565 different HLA-A, B, DR,DQ phenotypes (Table 1).

Typing methods

Since laboratories differ substantially with respect to the level of definition of broad and split alleles, this study concerns only the antigens shown in Table 2. Donors with different antigens were excluded from the analysis, that is, donors typed as A9 (broad antigen) and not split in A23 or A24. The 'blank' allele (bl) was allowed for each locus.

The HLA typing was performed according to the French Registry Quality Assurance Procedure, with HLA class I typing defined according to serological or molecular biology techniques. When the serological result was blank or uncertain, a DNA confirmatory typing was required. The HLA class I type was checked before the class II typing, which was performed with molecular biology techniques. The Transplant Centre performs confirmatory HLA class II typing when the donor is preselected for donation.

Surname methods

Surnames were used to specify the geographical origin of donors; origin was restricted, however, to one of the 90 French departments. $P_{jk,}$ the probability that phenotype j came from department k, was estimated as follows:

$$P_{jk} = \frac{\sum_{i=1}^{s} \omega_{ijk} f_{ik}}{\sum_{j=1}^{r} \sum_{i=1}^{s} \omega_{ijk} f_{ik}}$$

with

$$\omega_{ijk} = \frac{\delta_{ijk} \pi_{jk}}{\sum_{j=1}^{r} \delta_{ijk} \pi_{jk}}$$

with f_{ik} the frequency of the ith surname in the kth department. ω_{ijk} is a weight that takes into account the presence ($\delta_{ijk} = 1$) or absence ($\delta_{ijk} = 0$) of phenotype j in any person with surname i in department k and the *a priori* probability π_{jk} of phenotype j in department k. We summed over all surnames and standardised by the sum of the weights over all phenotypes.

Once the phenotype frequency P_{jk} is obtained, it can be replaced into the formula as a new estimate of the *a priori* probability. The final estimation of P_{jk} can be obtained by iterative calculation until its value converges. The purpose of this work, however, was to look for rare haplotypes, so we preferred to postulate that each phenotype has an equal *a priori* probability in each department. Accordingly, we wanted, insofar as possible, to overestimate very slightly the frequencies of rare phenotypes compared with those of common phenotypes.

By applying the formula, we could attribute to each of the 8558 HLA-A,B phenotypes and each of the 13 565

Table 1 Distribution of the numbers of donors and surnames by typed phenotypes

	HLA AB	*HLA ABDRDQ*
No. of donors	30 777	17 745
No. of surnames	17 971	11 667
No. of phenotypes	8558	13 565

Table 2 List of alleles studied

A	*B*	*DR*	*DQ*
1	7	2	1
2	8	3	2
3	13	4	3
11	18	7	4
23	22	8	
24	27	9	
25	35	10	
26	37	11	
28	38	12	
29	39	13	
30	41	14	
31	42		
32	44		
33	45		
34	47		
36	49		
66	50		
	51		
	52		
	53		
	57		
	58		
	60		
	61		
	62		
	63		
	64		
	65		
	67		
	70		
	75		
	77		
	78		

HLA-A,B,DR,DQ phenotypes (and thus to each donor) a vector of 90 probabilities of origin, one probability per department.

Statistical methods

We estimated the haplotype frequencies in each department, from the phenotypic frequencies provided by the method described above. These estimations were made with an expectation-maximisation algorithm, an iterative procedure that yields maximum-likelihood estimates of haplotype frequencies from multilocus genotype data with unknown gametic phase. This procedure is implemented in the 'Arlequin' software programme.[12] Hardy–Weinberg equilibrium of the HLA haplotypes cannot be calculated because of the presence of the recessive 'blank' allele.

The value for the minimum number of haplotypes (mnh) needed to reach the threshold of 0.50 frequency serves as a summary indicator of HLA geographical diversity in France.

Results

Distribution of HLA-A,B haplotypes

Most HLA-A,B haplotypes had a frequency lower than 0. 001 in all French departments: only 514 of 8558 haplotypes were present at a frequency estimated to be higher than 0.001. Table 3 lists the 63 HLA-A,B haplotypes found in all 90 departments. Since they represent the most widely shared genetic background in France, they are the least difficult to locate and obtain for transplantation.

Only 15 haplotypes were found in only one department. Most of these were present at a very low frequency (Table 4), except HLA-A3,B78 (0.003) in Landes (dept. no. 40) and HLA-A66,B50 (0.001) in Aube (dept. no. 10).

Table 5 summarises the most frequent haplotypes according to department. HLA-A1,B8 was the most common in three departments (dept. nos. 1, 59, 65) with its highest frequency (0.062) in Hautes-Pyrénées (dept. no. 65). HLA-A29,B44 was the most common in 50 departments, and its frequency was highest (0.078) in Landes (dept. no. 40). The frequency of the HLA-A2,B44 haplotype was highest (0.079) in Morbihan (dept. no. 56). The trend in Corsica (dept. no. 20) was different: HLA-A2,B62 was the most frequent haplotype (0.041) and HLA-A29,B44 the least frequent (0.005). Landes (dept. no. 40) and Tarn-et-Garonne (dept. no. 82) had their most frequent haplotype (HLA-A29,B12) in common, at a frequency of 0.062 and 0.052, respectively.

The mnh needed to reach 50% of the total frequency was approximately 25 and correlated with the homozygosity index (ie, the probability of finding donors with a homozygous genotype) but not with the total number of haplotypes in the department (data not shown).

The lowest mnh value (18) was observed in Hautes-Pyrénées (dept. no. 65) and the highest (31) in Meurthe-et-

Table 3 List of HLA-A, B haplotypes present in all 90 French departments

A	*B*
1	8
1	37
1	57
2	7
2	18
2	27
2	35
2	39
2	44
2	49
2	50
2	51
2	57
2	60
2	62
3	7
3	8
3	18
3	35
3	44
3	51
3	62
11	22
11	27
11	35
11	44
11	51
23	44
23	49
24	7
24	18
24	35
24	39
24	44
24	51
24	61
24	62
25	18
26	22
26	35
26	38
26	41
26	51
28	35
28	44
28	51
28	53
29	7
29	44
30	13
30	18
30	49
30	51
31	35
31	51
31	60
32	8
32	27
32	35
32	44
32	51
32	61
32	62

Moselle (dept. no. 54). The department of Paris (no. 75) also had a high mnh value (29), with all haplotypes present at a very low frequency, and HLA-A29,B44 the most frequent (0.05) (Figure1).

Distribution of HLA-A,B,DR,DQ haplotypes

In all, 5029 different HLA-A,B,DR,DQ haplotypes occurred at a frequency greater than 0.0001. Only 15 were found in all 90 departments at a frequency greater than 0.001 (Table 6). The most frequent haplotype was HLA-A1,B8,DR3,DQ2 (ranging from 0.077 in Morbihan to 0.012 in Vaucluse), followed by HLA-A29,B44,DR7,DQ2 and HLA-A3,B7,DR2,DQ1.

Substantial geographical diversity is observed in this haplotype distribution. The mnh necessary to reach a frequency of 0.5 ranged from 54 to 60 in 26 departments (Figure 2) and exceeded 60 in the others, except for Corsica (mnh only 20 haplotypes). Departments with a low mnh value were located in the Alps (dept no. 4), the Pyrénées (dept no. 9 and 65), and the Massif Central (dept. no. 29) and in Brittany (dept no. 56). Conversely, Paris, as a department, had the highest mnh value (ie, 97), each of haplotypes at a low frequency.

Table 7 reports the 14 haplotypes found at a frequency greater than 0.005 in only one department. Only two were observed at a frequency higher than 0.01.

Table 8 summarises the most frequent haplotypes in all 90 departments. HLA-A1,B8,DR3,DQ2 was found at a frequency of 0.077 in seven departments in the north of France, HLA-A29,B44,DR7,DQ2 was the most common haplotype in 53 departments (particularly in the south), and HLA-A3,B7,DR2,DQ1 in 28 departments (throughout the country). HLA-A2,B51,DR11,DQ3 was the most frequent (0.065) haplotype in Corsica and HLA-A2,B7,DR2,DQ1 (0.068) in Hautes-Pyrenees.

Table 4 List of HLA-A, B haplotypes found in only a single department (Dept. no.) and their frequency

A	*B*	*Dept. no.*	*Frequency*
1	77	89	0.00020
2	42	43	0.00028
3	75	88	0.00016
3	78	40	**0.00317**
25	47	86	0.00063
30	42	85	0.00085
30	67	1	0.00023
31	41	48	0.00014
33	78	76	0.00017
34	22	55	0.00012
66	44	56	0.00068
66	47	44	0.00064
66	50	10	**0.00159**
66	60	62	0.00013
66	70	45	0.00013

Discussion

Surnames have proved to be useful for describing the population structure, and isonymy makes it possible to evaluate isolation and consanguinity (see the recent bibliography in Barrai *et al*,[13] Mathias *et al*,[14] Rodriguez-Larralde,[15] Lasker *et al*,[16] Sykes *et al*,[17] Manni *et al*,[18] Barrai *et al*,[19,20] Caravello *et al*,[21] Gagnon *et al*,[22]). They are also frequently used as a tool to identify individuals' geographical origin (see the recent bibliography in Chan,[23] Garza-Chapa *et al*,[24] De Silvesti and Guglielmino,[25] Pollock *et al*,[26] Jacobs and Landerdale,[27] Rudan,[28] Carta *et al*,[29] Brancatelli *et al*[30]). Surname distribution in France has been accurately known since at least the end of the 19 century and has been studied extensively.[31–42]

Under these circumstances, the use of surnames might provide additional specificity about the geographical origin of bone marrow donors, and, consequently, increase our efficiency in locating wanted donors. First, we can easily recognise donors who have migrated to France relatively recently: their surnames are not registered in any department before 1916. This point indicates the essential arbitrariness of our definition of the 'French population'

Table 5 List of the most frequent HLA-A, B haplotypes in France, estimated as the mean of the frequency in each of the 90 departments. The minimum and maximum frequencies are given with the department where they occur (Depmin and Depmax, respectively)

	France	*Min*	*Depmin*	*Max*	*Depmax*
A1B8	0.037	0.015	5	0.062	65
A2B7	0.020	0.005	12	0.039	29
A2B35	0.012	0.001	48	0.042	40
A2B44	0.047	0.029	46	0.079	56
A2B51	0.029	0.012	22	0.058	19
A2B62	0.022	0.008	68	0.041	20
A3B7	0.042	0.020	40	0.073	29
A3B35	0.023	0.008	40	0.041	9
A11B35	0.019	0.009	40	0.038	3
A24B35	0.014	0.003	48	0.037	20
A28B44	0.009	0.003	2	0.037	19
A29B44	0.053	0.005	20	0.078	40
A30B18	0.014	0.003	22	0.060	65

Table 6 List of the HLA-A, B,DR,DQ haplotypes present in all 90 French departments

A	*B*	*DR*	*DQ*
1	8	3	2
2	7	2	1
2	18	3	2
2	44	4	3
2	44	13	1
2	51	1	1
3	7	2	1
3	35	1	1
11	35	1	1
23	44	7	2
24	62	13	1
25	18	2	1
29	44	7	2
30	13	7	2
30	18	3	2

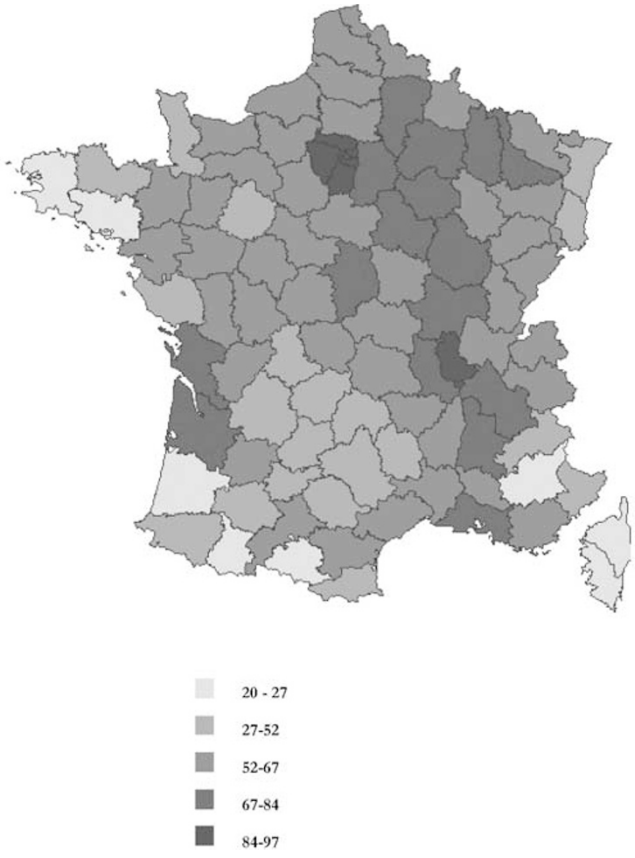

Figure 2 Departmental distribution of minimum number of HLA-A, B,DR,DQ haplotypes needed to reach 50% of the total frequency.

as people sharing names with people born and registered anywhere in France before 1916. France has traditionally welcomed emigrants from neighbouring countries (Italy, Spain, Belgium, etc), and, more recently, from Poland,

Table 7 List of the HLA-A,B,DR,DQ haplotypes found in an unique department (Dept. no.) with a frequency greater than 0.005

A	*B*	*DR*	*DQ*	*Dept. no.*	*Frequency*
2	7	10	3	46	0.00748
2	18	4	1	29	0.00696
2	41	3	3	20	0.01257
57	7	1	20	2	0.00622
3	44	8	1	5	0.00501
3	47	1	1	65	0.00567
23	35	8	3	74	0.00839
24	22	1	3	29	0.00579
26	35	4	2	20	0.01268
28	7	10	3	12	0.00677
29	64	13	1	12	0.00677
31	64	4	3	67	0.00651
33	8	3	2	9	0.00534
33	49	8	2	40	0.00613

Armenia, North Africa and Asia. 'The French' cannot therefore be realistically considered as a homogeneous genetic unit. Various studies[9,32,43] have already emphasised the large genetic variability in France, which results from isolation and genetic drift in some areas and from large melting pots in others. Even these studies, however, are quite imprecise in their description of the French population structure. As Cavalli-Sforza *et al*[44] said 'Unfortunately, for a vast territory like France, 14 points do not permit a high resolution …'. The method described here, using surname data, allows us to circumvent this difficulty and provide an accurate geographical picture of the French genetic structure. Our principal purpose must nevertheless be borne in mind: to obtain an accurate geographical distribution of HLA haplotypes to optimise policies for the recruitment of bone marrow donors.

HLA-A,B haplotypes

We have not focused on the most frequent HLA-A,B haplotypes, found all over France (Table 3). These are usually not specific to France and are found, for example, distributed at high frequencies in other western European countries.[45] Table 4 provides more useful results inasmuch as these 'unique' and rare haplotypes are more difficult to locate when needed. They are either typical haplotypes attached to a given department or foreign haplotypes from recent immigration. Not surprisingly, the probability of findings these kinds of haplotypes is high in Paris (dept no. 75) and its suburbs (dept no. 77 and 78), as a consequence of the convergent migration towards Paris from all departments, particularly since the 19th century. Indeed, almost all haplotypes (378 of the 514 with a frequency larger than 0.01) can be found in department 75, albeit at a very low frequency: the mnh (that is, the number needed to reach a frequency of 0.50) is 29. The most common haplotype in Paris is HLA-A29,B44 (0.05), which is also found throughout western Europe, in Finland, and even

Table 8 List of the most frequent HLA-A, B,DR,DQ haplotypes in France, estimated as the mean of the frequency in each of the 90 departments frequencies. The minimum and maximum frequencies are given with the department where they occur (Depmin and Depmax, respectively)

	France	*Min*	*Depmin*	*Max*	*Depmax*
A1B8DR3DQ2	0.032	0.012	20	0.077	56
A2B44DR4DQ3	0.023	0.008	68	0.059	29
A2B44DR7DQ2	0.008	<0.001	4	0.043	9
A2B44DR13DQ1	0.013	<0.001	22	0.037	12
A2B51DR11DQ3	0.008	<0.001	56	0.065	20
A3B7DR2DQ1	0.036	0.016	20	0.075	29
A3B35DR1DQ1	0.015	0.002	5	0.034	64
A28B35DR8DQ4	0.001	<0.001	30	0.046	65
A29B44DR7DQ2	0.042	0.015	20	0.087	40

among the Zulus of sub-Saharan Africa.[45] From a transplantation point of view, this means that theoretically all HLA-A,B haplotypes can be found in Paris, but at so low a frequency that without high effective recruitment policies (for example, completion of a questionnaire about geographical origin even before typing analysis), it is nearly impossible to find a specific rare haplotype.

Most of the departments in France have different and original patterns (Figure 1). Departments that are made up of scattered villages due to hilly landscapes or hedged fields (dept. nos. 4, 9, 12, 29, 56, 65, 81) have the fewest haplotypes and lowest mnh. The rarity of immigrants contributed to the high degree of consanguinity.[31] Conversely, the departments that welcomed many immigrants have the most variant haplotypes (Lorraine, dept. nos. 8, 54, 57; Ile-de-France, dept. nos. 75, 77, 10, 89; Provence, dept. nos. 30, 13). One particular example is Corsica, which has a long history of isolation within the Mediterranean basin, despite a close relation with Italy. It was annexed to France in 1768. Its mnh is only 20 haplotypes. The two most frequent haplotypes there are also found in western Europe, in particular in Spain and Italy (HLA-A2,B62), and in Spain, Portugal and Greece (HLA-A24,B35).[45] Many haplotypes are also shared with the nearest mainland departments (depts. nos. 6, 83 and 13), where recent emigrants from Corsica first settled.

HLA-A,B,DR,DQ haplotypes

The frequencies of the HLA-A,B,DR,DQ haplotypes were estimated from fewer people (17 745 donors) than those frequencies for the HLA-A,B haplotypes (30 777 donors) because until the end of 1998, donors were typed for DR and DQ only when their potential HLA-A, B was compatible with a recipient. This could introduce a bias into the estimation of haplotype frequencies. Thereafter donors were simultaneously typed for HLA-A, B, DR, DQ, but unfortunately this sample is still much too small (2216) for a detailed description of all of France.

As reported above, Corsica is different from the rest of the country, with fewer HLA-A, B, DR, DQ haplotypes at higher frequencies. Figure 2 shows that again only 20 of the most frequent haplotypes was sufficient to reach the value of 0.5: Corsica has less variability than the other departments. Table 7 indicates that of the haplotypes found in only a single department, two were located in Corsica at a frequency greater than 0.01: HLA-A26,B35,DR4,DQ2 and HLA- A2,B41,DR3,DQ3. The latter haplotype contains HLA-B41, also found in Lebanon, Ukraine and Armenia,[45] and HLA-A2,B41, which has a high frequency in Greece.[45] The presence of these haplotypes in Corsica suggests the possibility of genetic exchange or migration between these regions and Corsica.

The frequency of the HLA-A28,B35,DR8,DQ4 haplotype shows the greatest contrast. It seems to have originated in the Pyrenees; its frequency decreases rapidly as it spreads towards neighbouring departments, and it is absent from most other departments (55 of them). This distribution may reflect the trace of the 'Basque' population living in France and Spain.

Although the frequencies of the HLA-A,B,DR,DQ haplotypes are slightly biased, as explained above, their departmental distributions are consistent with those of the HLA-A,B frequencies,

This detailed description of the HLA diversity in France uses a method that takes advantage of the known surname distribution to refine the results described by Lonjou *et al*[9] and to increase the accuracy of the geographical origin of haplotypes. Doing so enhances the likelihood of localising and obtaining, at the lowest cost, the best donors for patients awaiting bone donor transplantation from unrelated donors, particularly when their haplotypes are rare: first by identifying the geographical origin of the intended recipient, and then, by using the method described here to locate the area where the probability of finding the volunteer donors sharing these same haplotypes is highest.

Acknowledgements

This work was supported by Grant EFG 1999 and by grant of OHLL CNRS program. We thank all the French typing laboratories and the French donors for their kindness.

References

1 Darke C: An overview of the Welsh bone marrow donor registry 10 years of bone marrow donor provision. *Bone Marrow Transplant* 2000; **25**: 771–777.
2 Dini G, Cancedda R, Locatelli F *et al*: Unrelated donor marrow transplantation: an update of the experience of the Italian Bone Marrow Transplant Group (GITMO). *Haematologica* 2000; **85**: 30–36.
3 Ordemann R, Petzold K, Holig K *et al*: The Dresden Cord Blood Bank. Experiences of the Cord Blood Bank in Dresden, promoted by the German bone Marrow Donor Registry *Dtsch Med Wochenschr* 2000; **125**: 1424–1428.
4 Velickovic ZM, Carter JM: Feasibility of finding an unrelated bone marrow donor on international registries for New Zealand patients. *Bone Marrow Transplant* 1999; **23**: 291–294.
5 Gyodi E, Rajczy K, Penzes M *et al*: Hungarian Bone Marrow Donor Registry: present status and activity. *Bone Marrow Transplant*. 1998; **22**: S38–S40.
6 Kupkova L, Ivaskova E, Hana I *et al*: An overview of the Czech Bone Marrow Donor Registry. *Bone Marrow Transplant*. 1998; **22**: S41–S44.
7 Widmer-Kuhnlein E, Gratwohl A, Kern M *et al*: 10 years Swiss bone marrow donor registry. *Schweiz Med Wochenschr.* 1998; **128**: 1602–1604.
8 Martinetti M, Degioanni A, D'Aronzo AM *et al*: An immunogenetic map of Lombardy (Northern Italy). *Ann Hum Genet* 2002; **66**: 37–48.
9 Lonjou C, Clayton J, Cambon-Thomsen A, Raffoux C: HLA-A, B, DR haplotypes in France: implications for recruitment of potential bone marrow donors. *Transplantation* 1995; **60**: 375–383.
10 Crognier E, Bley D, Boetsch G: *Mariage en Limousin, Evolution séculaire et idéntité d'une population rurale, le canton de Chateauponsac (1870–1979)*. Paris: CNRS Editions, 1998.
11 Darlu P, Degioanni A, Ruffié J: Quelques statistiques sur la distribution des patronymes en France. *Population* 1997; **3**: 607–634.
12 Schneider S, Kueffer JM, Excoffier L: *Arlequin (Version 2000) a Software for Population Genetic Data Analysis*. Switzerland: Genetics and Biometry Laboratory University of Geneva, 1997.
13 Barrai I, Rodriguez-Larralde A, Mamolini E, Manni F, Scapoli C: Elements of the surname structure of Austria. *Ann Hum Biol* 2000; **27**: 607–622.
14 Mathias RA, Bickel CA, Beaty TH *et al*: A study of contemporary levels and temporal trends in inbreeding in the Tangier Island, Virginia, population using pedigree data and isonymy. *Am J Phy Anthropol* 2000; **112**: 29–38.
15 Rodriguez-Larralde A, Morales J, Barrai I: Surname frequency and the isonymy structure of Venezuela. *Am J Hum Biol* 2000; **12**: 352–362.
16 Lasker GW, Mascie-Taylor CGN: The genetic structure of English villages: surname diversity changes between 1976 and 1997. *Ann Hum Biol* 2001; **28**: 546–553.
17 Sykes B, Irven C: Surnames and the Y chromosome. *Am J Hum Genet* 2000; **66**: 1417–1419.
18 Manni F, Barrai I: Genetic structures and linguistic boundaries in Italy: a microregional approach. *Hum Biol* 2001; **73**: 335–347.
19 Barrai I, Rodriguez-Larralde A, Mamolini E, Manni F, Scapoli C: Isonymy structure of USA population. *Am J Phy Anthropol* 2001; **114**: 109–123.
20 Barrai I, Rodriguez-Larralde A, Manni F, Scapoli C: Isonymy and isolation by distance in the Netherlands. *Hum Biol* 2002; **74**: 263–283.
21 Caravello G, Tasso M: Use of surnames for a demo-ecological analysis: a study in southwest Sardinia. *Am J Hum Biol* 2002; **14**: 391–397.
22 Gagnon A, Toupance B: Testing isonymy with paternal and maternal lineages in the early Quebec population: the impact of polyphyletism and demographic differentials. *Am J Phy Anthropol* 2002; **117**: 334–341.
23 Chan C: The quality of life of women of Chinese origin. *Health Soc Care Community* 2000; **8**: 212–222.
24 Garza-Chapa R, De Los Angeles Rojas-Alvarado M, Cerda-Flores RM: Prevalence of NIDDM in Mexicans with paraphyletic and polyphyletic surnames. *Am J Hum Biol* 2000; **12**: 721–728.
25 De Silvestri A, Guglielmino CR: Ethnicity and malaria affect surname distribution in Consenza Province (Italy). *Hum Biol* 2000; **72**: 573–583.
26 Pollock BH, DeBaun MR, Camitta BM *et al*: Racial differences in the survival of childhood B-precursor acute lymphoblastic leukemia a Pediatric Oncology Group Study. *J Clin Oncol* 2000; **18**: 813–823.
27 Jacobs EA, Lauderdale DS: Receipt of cancer screening procedures among Hispanic and non-Hispanic health maintenance organization members. *Cancer* 2001; **91**: 257–261.
28 Rudan I: Ancestral kinship and cancer in Lastovo Island, Croatia. *Hum Biol* 2001; **73**: 871–884.
29 Carta MG, Kovess V, Hardoy MC, Morosini P, Murgia S, Carpiniello B: Psychiatric disorders in Sardinian immigrants to Paris: a comparison with Parisians and Sardinians resident in Sardinia. *Soc Psychiatry Psychiatr Epidemiol* 2002; **37**: 112–117.
30 Brancatelli G, Federle MP, Grazioli L, Carr BI: Hepatocellular carcinoma in noncirrhotic liver: CT, clinical, and pathologic findings in 39 US residents. *Radiology* 2002; **222**: 89–94.
31 Darlu P, Ruffie J: Relationships between consanguinity and migration rate from surname distributions and isonymy in France. *Ann Hum Biol* 1992; **19**: 133–137.
32 Degioanni A, Lisa A, Zei G, Darlu P: Patronymes italiens et migration italienne en France entre 1981 et 1940. *Population* 1996; **6**: 1153–1180.
33 Darlu P, Degioanni A, Ruffié J: Quelques statistiques sur la distribution des patronymes en France. *Population* 1997; **3**: 607–634.
34 Legay JM, Vernay M: The distribution and geographical origin of some French surnames. *Ann Hum Biol* 2000; **27**: 587–605.
35 Vernay M: Trends in inbreeding, isonymy, and repeated pairs of surnames in the Valserine Valley, French Jura, 1763–1972. *Hum Biol* 2000; **72**: 675–692.
36 Mourrieras B, Darlu P, Hochez J, Hazout S: Surname distribution in France: a distance analysis by a distorted geographical map. *Ann Hum Biol* 1995; **22**: 183–198.
37 Degioanni A, Darlu P: A Bayesian approach to infer geographical origins of migrants through surnames. *Ann Hum Biol* 2001; **28**: 537–545.
38 Darlu P, Degioanni A, Jakobi L: Les cloisonnements dans les Pyrénées occidentales. Evolution du XIX siècle à nos jours. In Brunet G, Darlu P, Zei G (eds) *Le patronyme. Histoire, anthropologie, société*. Paris: CNRS Editions, 2001, pp 173–188.
39 Degioanni A, Darlu P, Poulain M, Foulon M: Immigration flamande en France au XIX siècle et au débout du XX siècle. In Brunet G, Darlu P, Zei G (eds): *Le patronyme. Histoire, anthropologie, société*. Paris: CNRS Editions, 2001, pp 189–203.
40 Boetsch G, Prost M, Rabino-Massa E: Evolution of consanguinity in a French alpine valley: the Vallouise in the Brianfon region (17^{th}–19th Centuries). *Hum Biol* 2002; **74**: 285–300.
41 Morelli L, Paoli G, Francalacci P: Surname analysis of the Corsican population reveals an agreement with geographical and linguistic structure. *J Biosociol Sci* 2002; **34**: 289–301.
42 Degioanni A, Lisa A, Zei G, Darlu P: Patronymes italiens et migration italienne en France entre 1981 et 1940. *Population* 1996; **6**: 1153–1180.
43 Ohayon E, Cambon-Thomsen A: *Génétique des Population Humaines*, Paris, INSERM 1986.
44 Cavalli-Sforza LL, Menozzi P, Piazza A: *The history and geography of human genes*. Englewood Cliffs NJ Princeton University Press, 1994.
45 Charron D: *HLA Genetic Diversity of HLA Functional and Medical Implication*. EDK Paris: Dominique Charron, 1997.