Practical Application of Population Genetics : the Genetic Survey "Provinces Françaises"

A. Cambon-Thomsen*, E. Ohayon*
and the Participants of the Genetic Survey "Provinces Françaises"**

* INSERM U 100, CHU Purpan, 31052 Toulouse cedex, France

INTRODUCTION

From 1981 to 1985, a genetic survey in the French provinces and in Quebec was conducted by more than 20 laboratories specialized in immunogenetics and human genetic polymorphism studies, under the coordination of E. Ohayon and A. Cambon-Thomsen. A large panel of genetic markers was studied in 1382 families of precisely known French origin ; the data bank was completed and population genetics analyses were performed between 1985 and 1987.

In this report we will describe the bases of this study, the methodology of the survey with comments about the samples of population studied, the organization of the data bank "PF" (for "Provinces Françaises") and the methods used for the population genetics analysis. Some of the results will be presented but more details about this collaborative work may be found in several volumes published by INSERM, Paris (Ohayon and Cambon-Thomsen 1986, 1987).

**
- Participants in Data Analysis : N. Borot, A. Sevin, Centre de Recherches sur le Polymorphisme Génétique des populations Humaines (CRPG), CNRS, Toulouse ; M. Neugebauer, M. Baur, Institute for Medical Statistics, Bonn ; G. Sierp, E. Albert, Immunogenetics Laboratory, Munich ; M. Thomsen, INSERM U 100, Toulouse.
- Participants in Survey and Regional Data Production : F. Dufossé et al (Lille), M.M. Tongio et al (Strasbourg), P. Perrier et al (Nancy), H. Bétuel et al (Lyon), F. Guignier et al (Dijon), J.C. Bensa et al (Grenoble), P. Mercier et al (Marseille), J.P. Amoros et al (Ajaccio), J. Seignalet et al (Montpellier), J.Y. Muller et al (Paris), Ph. Lefèvre-Witier et al (Toulouse), J. Hors et al (Paris), M. Calot et al (Toulouse), G. Malinvaud et al (Limoges), C. Moennard et al (Clermont-Ferrand), D. Alcalay et al (Poitiers), R. Fauchet et al (Rennes), J.P. Saleun et al (Brest), A.M. Griveau et al (Caen), F. Décary et al (Montréal, Canada).
- Participants in Centralized Studies for Some Markers : J. Arnaud, M. Blanc, J. Constans, CRPG-CNRS, Toulouse, M. Abbal, INSERM U 100, Toulouse, J. Ducos, CTS, Toulouse, G. Hauptmann, CTS, Strasbourg.

Advances in Forensic Haemogenetics 2
Edited by W.R. Mayr

BASES OF THE COLLABORATIVE STUDY "PROVINCES FRANCAISES"

Since Landsteiner discovered the first blood group system (ABO) in 1900, the study of genetic markers has become a large part of human population genetics. Variable gene frequencies were found in different populations and genetic distances could be calculated and compared to geographical data (Cavalli-Sforza and Bodmer 1971 ; Mourant 1985 ; Piazza and Menozzi 1983 ; Salmon et al 1984). However new interests in such studies arose about 15 years ago as **1)** new genetic systems with higher polymorphism were discovered, **2)** some of them were shown to have a fundamental biological role, such as the HLA system regarding immune response and **3)** genetic epidemiology began to be recognized as important and brought population genetics to the field of medicine (Degos et al 1977 ; Morton 1982).

It is well known that the French population has a heterogeneous origin and a complicated history and previous studies have already demonstrated that genetic markers are not evenly distributed between regions (Mayer et al 1981 ; de Mouzon et al 1980 ; Prévost et al 1984 ; Vergnes et al 1980 ; Youinou et al 1983). However there were no homogeneously established data in France on the distribution of genetic markers in different regions, based on family studies. We therefore undertook this work in order **1)** to describe the distribution of gene frequencies in the French provinces, **2)** to organize a data bank which could be useful in different fields (human genetics, anthropology, history, demography, medicine and epidemiology), **3)** to produce reference maps of gene frequencies **4)** to allow the study of inter-regional variability in France on the basis of genetic data.

METHODOLOGY OF THE SURVEY "PROVINCES FRANCAISES"

The originality of this study mainly consists in the methodology : it is a prospective study, with homogeneous methods (within the limits imposed by a multicenter study) both for the choice of the samples and for the reagents and techniques used. The principal steps were the following :

1. Call for Participants :

This was done in 1980 and 20 laboratories were willing to participate. They became a determining element in defining the size, the location and the number of population samples which could be studied.

2. Choice of Areas :

The rules were to cover as large a part of France as possible, to avoid migration axes and zones of known population admixture as well as isolates and to take into account

geographical and historical criteria. Fifteen French provinces were chosen and a sample of Canadians of French origin from Quebec was also included in the study. These provinces designated by a code letter were studied by the different volunteer laboratories under the responsibility of a regional coordinator (name between brackets) : F, Flanders (F. Dufossé), J1, Alsace (M.M. Tongio), J2, Lorraine (P. Perrier), B, Beaujolais-Burgundy (H. Bétuel, F. Guignier), D, Savoy-Dauphiné (J.C. Bensa), V, Corsica (J.P. Amoros, P. Mercier), H, Cévennes (J. Constans, J.Y. Muller, J. Seignalet), C, Catalonia (J. Hors, Ph. Lefèvre-Witier), G, Béarn (M. Calot), L, Limousin (G. Malinvaud), A, Auvergne (C. Moennard), P, Poitou (D. Alcalay), S, South Brittany (R. Fauchet), N, North Brittany (J.P. Saleun), T, Low Normandy (A.M. Griveau), Q, Quebec (F. Décary). Each regional coordinator was responsible for defining the precise geographical limits of the province, which could be divided in "sub- regions", and for finding the families homogeneously distributed over the whole province.

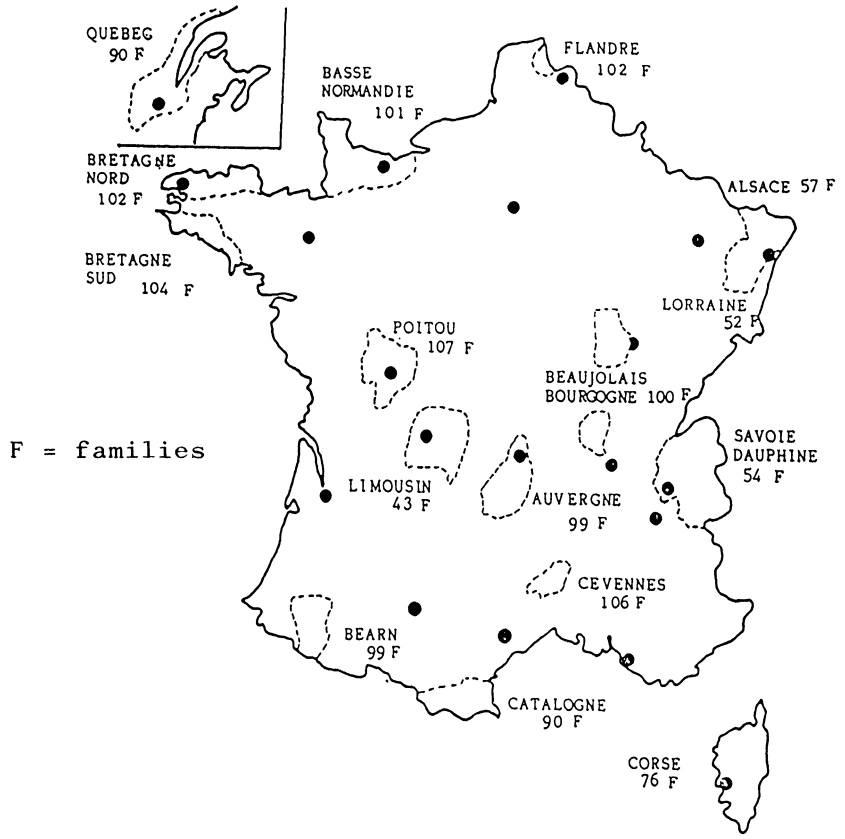

Fig 1. Distribution of the families studied in the genetic survey "Provinces Françaises". The names of the provinces are given in French. The total number of families is 1382; 1360 of them could be fully analyzed.

3. Choice of Families :

The rule was to choose families who had been settled in the province for at least three generations. For this reason most of the families lived in rural areas where the migration rate is low. Each family should be available for answering a questionnaire and for giving a blood sample from at least the 2 parents and 2 children over 10 years. Within each province the families were contacted with the help of blood banks, schools, articles in local journals, local administrators, social organizations and also through the families previously contacted especially for completing the sample at the end of the survey.

4. Sample Size :

The number of 100 unrelated families per province was chosen after calculating the probability not to observe an allele of frequency 1% according to the sample size. This probability is between 1 and 2% for a sample of 200 individuals. Figure 1 presents the distribution of the 1382 families among the provinces studied. The analysis could be done for 1360 of them, which represents a total of 5500 persons with 2720 unrelated individuals. An important fact is that the families were selected according to given criteria and, consequently, were not random. The total sample (all provinces together) is not intended to represent "the" French population and this study was mainly done for the comparison of samples of approximatively the same size. Fourty-three to 106 families were studied per province. The low number of families in Limousin (43), Savoy-Dauphiné (54) and Corsica (76) implies that the results in terms of frequencies should be considered with caution. At the beginning of the study Alsace and Lorraine were put together with around 50 families in each. But at the time of the analysis we considered them independently because a strong genetic heterogeneity appeared. The North-Western part of France is slightly overrepresented : 414 families, as compared to 232 from the South-Western part, 311 from the North-East and 335 from the South-East). This is also explained by the fact that some regions like Aquitaine (South-West) adn Provence (South-East) which would have been interesting to study were not considered because none of the laboratories could perform the survey and the typing for these provinces. Thus our sample may be representative of a stable rural population of some French provinces at the beginning of this century ; but one of the major interests is to provide us with genetic data on a large sample of families.

5. Panel of Genetic Markers :

The following genetic systems were studied and we indicate between brackets the chromosome (ch) where the corresponding loci are located :

Erythrocyte blood groups :

Typing for ABO (ch 9), Rhesus (ch 1), MNSs (ch 4), P (ch 22) and Kell was done by classical haemagglutination methods in the laboratories of each province.

Major histocompatibility complex (MHC) : the human MHC is the most polymorphic genetic system with polyallelic multiple loci and is very informative in population genetics (Baur and Danilovs 1980 ; Greenacre and Degos 1977 ; Ryder et al 1978). It is located on the chromosome 6 and includes HLA genes (HLA-A,B,C,DR,DQ,DP) and non HLA genes (some complement genes, 21-hydroxylase genes, etc...). An erythrocyte enzyme, the glyoxalase-I (GLO) is also coded nearby this region. The HLA antigens HLA-A,B,C and DR were determined by lymphocytotoxicity in the regional laboratories with the help of a common set of 192 alloantisera previously selected by 8 of the laboratories. The complement allotypes BF, C4A and C4B were studied by electrophoresis and immunofixation mainly in two laboratories (G. Hauptmann, Strasbourg and M. Abbal, Toulouse) although 5 other laboratories also performed this typing for some provinces. Finally GLO was studied by electrophoresis principally by J. Arnaud (Toulouse) and M. North (Strasbourg), and samples from 3 provinces were studied in other laboratories.

Serum proteins

All the following systems were studied centrally for all provinces in one laboratory in Toulouse, Centre de Recherches sur le Polymorphisme Génétique des Populations Humaines (CRPG), CNRS :

Immunoglobulin allotypes of systems Gm (ch 14) and Km (ch 2) were determined by inhibition of haemagglutination (M. Blanc).

Serum proteins with electrophoretic polymorphism were investigated by J. Constans : Protein inhibitor (Pi) or alpha$_1$ antitrypsin (ch 14), Group Specific Component (Gc) or vitamin D binding protein (DBP) (ch 4), haptoglobin (Hp) (ch 16) and transferrin (Tf) (ch 3). Pseudocholinesterase polymorphism (ch 1) was tested by J. Arnaud.

A detailed description of these genetic systems, the distribution of their alleles in populations and the detailed results in PF may be found in Ohayon and Cambon-Thomsen (1986, 1987).

6. Data Entry and Control of Data Set :

The data entry was centralized by A. Sevin (CNRS, Toulouse). The common coding forms allowed to create two kinds of data files : **1)** a data set per family, based on the answers to a questionnaire giving origin, location, identification (number) and pedigree of the family along with some socio- cultural and medical information ; **2)** a data set of laboratory results per individual with HLA serological reaction scores and local

The control of data and data entry was done at two levels : **1)** by the regional coordinator in each province, upon receipt of a clear listing of the data entered ; **2)** by the centralizing team using both computer programs and human ability to correct the considerable amount of logical and clerical errors which are unavoidable when data files consist of contributions from numerous laboratories.

DATA ANALYSIS

This analysis was performed in several steps before the data bank was made available for the scientific community : the final data set and the estimation of the population genetics parameters are the fruit of an international collaboration between INSERM U 100, Toulouse ; CRPG-CNRS, Toulouse ; Institute of Medical Statistics, Bonn and Immunogenetics Laboratory, Munich.

1. HLA Serological Analysis

The programs were derived from those used in 1984 for the 9th international histocompatibility workshop (Albert et al 1984, Deppe et al 1984). Based on the serological reaction scores from the 192 common alloantisera and on the laboratory HLA antigen assignment they allowed to evaluate the quality of the serums (Q scores), the quality of antigen definition (Table I) and to perform a computer cell typing. The overall correlation coefficient between laboratory and computer antigen assignment was equal to 0.93. After this analysis a new set of HLA phenotypes was created for all individuals tested for further genetic analyses.

2. Population Genetics Analysis :

Two approaches were used for the estimation of gene frequencies in each province and for the total sample :

- without considering the family information, the antigen and phenotype frequencies were calculated from the unrelated individuals (parents) by a simple counting method and the gene frequencies were estimated by maximum likelihood ; for HLA genes the formula $f = 1-\sqrt{1-F}$ where f is the gene frequency and F the antigen frequency was used.

- with family analysis, using the set of programs FAP originally written for the 9th international histocompatibility workshop in 1984 and adapted for this analysis (Neugebauer et al 1984, Borot et al 1986). This program was first used for analyzing the MHC markers on **a)** data from laboratory HLA antigen assignment, **b)** corrected data after computer serological analysis ; then it was adapted for application to all the other systems and this analysis is currently being completed. This program is excellent for checking family data consistency in large data sets ; it estimates the likelihood of all

possible genotypes or haplotype combinations compatible with the given phenotypes in each family ; it detects possible recombinations. It is therefore extremely helpful for analyzing multiple loci systems : it allowed to estimate the frequency of 8 locus haplotypes : HLA-A,C,B,BF,C4A,C4B,DR,GLO, as well as allele frequencies, to estimate the recombination fraction and to calculate 2 locus linkage disequilibrium.

- Hardy-Weinberg equilibrium was tested for all systems.

3. Inter-Regional Analysis

As the gene frequencies showed strong differences between provinces a number of inter-regional analyses were undertaken. We will only show here results of genetic distances calculated according to Balakrishnan and Sanghvi (1968). Other indexes of genetic distances have been used (Cambon-Thomsen et al 1986) and a principal component analysis leading to generate coloured synthetic maps was performed by A. Piazza (1986).

SOME RESULTS OF THE PF ANALYSIS :

Table 1 shows the quality of HLA antigen definition as result of the serological analysis.

Table 2 gives the gene frequencies of non MHC markers estimated by maximum likelihood from the phenotypes of the parents, on the total sample.

Table 3 compares the HLA allele frequencies obtained by different methods from the total data set : (I) calculated from the antigen frequencies and based upon local antigen assignment, (II) estimated by FAP from the same data, (III) estimated by FAP using the corrected data set after serological analysis by computer ; allele frequencies for BF, C4A, C4B and GLO were also estimated by FAP.

Table 4 shows the most frequent HLA haplotypes in the overall sample by order of decreasing frequencies.

Figure 2 gives an example of an HLA-allele distribution (HLA-DR4) : there are strong differences between provinces with frequencies (%) decreasing according to a North-West/ South axis.

Figure 3 shows three dendrograms drawn from the matrix of X^2 genetic distances for three loci (ABO, HLA-B and HLA-DR).

Figure 4 illustrates the genetic resemblances between French provinces derived from the matrix of genetic distances calculated for a group of independent genetic systems (ABO, Rh, P, Kell, HLA-B, Gc, Pi, Tf). The number of lines between provinces is maximum when the distance is minimum.

COMMENTS AND CONCLUSION

The FAP program gave the most accurate estimation of gene and haplotype frequencies. However as shown in Table 3 the advantage of using a family analysis program is not so much for allele frequencies estimations, which vary only slightly according to the methods used ; it is mainly for haplotype and linkage analyses. This family material allowed to consider extended HLA haplotypes and complotypes ; the analysis of recombinations gave recombination fraction values (ϑ) concordant with those of the litterature : HLA-A/C : 0.47 and 0.95, HLA-C/B : 0.06 and 0.09, HLA-B/BF : 0.47 and 0.61, HLA-DR/BF : 1.13 and 0.94, HLA-DR/GLO : 5.72 and 11.16 for paternal and maternal ϑ values respectively.

When performing inter-regional comparisons, a remarkable genetic heterogeneity was found and allowed to define regional genetic characteristics. The sample from Corsica was very different from all the others ; although the general picture of gene frequency gradients for individual genetic systems is quite complex a clear difference between the provinces of Northern France and those of Southern France is constantly found.

Numerous other analyses will be done in the future as this study has produced one of the largest available data set on genetic markers based on a family material homogeneously studied. But one must be careful not to overinterpret the results in terms of history or population migrations.

REFERENCES

Albert ED, Barbalho T, Baur MP, Christiansen FT, Deppe H, Keller E, Louton T, Marshall WH, McNicholas A, Raffoux C Schießl B, Scholz S, Sierp G, Sigmund M (1984) The serological analysis of the data of the ninth histocompatibility workshop. In: Histocompatibility testing, Albert ED, Baur MP, Mayr WR (eds), Springer-Verlag, Heidelberg, 72-115

Balakrishnan V, Sanghvi LD (1968) Distance between populations on the basis of attribute data. Biometrics 24 859-865

Baur MP, Danilovs JA (1980) Population analysis of HLA-A, B, C, DR and other genetic markers. In: Histocompatibility testing, Terasaki PI (ed), UCLA, Los Angeles, pp 955-993

Borot N, Neugebauer M, Sevin A, Cambon-Thomsen A, Hillat L (1986) Méthode d'analyse génétique des données familiales dans l'enquête "Provinces Françaises". In: Génétique des populations humaines, Ohayon E, Cambon-Thomsen A, (Eds), Colloque INSERM vol. 142, pp 169-176

Cambon-Thomsen A, Borot N, Sevin A (1986) Eléments d'analyse inter-régionale : distances génétiques entre les provinces françaises. In: Génétique des populations humaines, Ohayon E, Cambon-Thomsen A, (Eds), Colloque INSERM vol. 142, pp 323-338

Cavalli-Sforza LL, Bodmer WF (1971) The genetics of human populations, Freeman, San Francisco
Degos L, Chaventré A, Jacquard A (1977) Migration, sélection et système d'histocompatibilité. Population 32 : 123-136
Deppe H, Albert ED, Baur MP (1984) Cell typing by computer. In: Histocompatibility testing, Albert ED, Baur MP, Mayr WR (eds), Springer-Verlag, Heidelberg, pp 60-63
Greenacre MJ, Degos L (1977) Correspondence analysis of HLA gene frequency data from 124 population samples. Am. J. Hum. Genet. 29 : 60-75
Mayer S, Grangé D, Tongio MM, Hauptmann G, Klein J (1981) Haplotype study of the HLA-A, B, C and Bf associations by the factorial correspondence analysis. Journal of Immunogenetics 8 : 7-13
Morton NE (1982) Outline of genetic epidemiology, S. Karger, Basel
Mourant AE (1985) Blood relations. Blood groups and anthropology. Oxford University Press (ed), Oxford
Mouzon (de) A, Constans J, Sommer E, Sevin A, Quilici JC, Fernet P, Ducos J, Ohayon E (1980) HLA-A, B typing in Basque and other pyrenean populations. Tissue Antigens 15 : 11-18
Neugebauer M, Willems J, Baur MP (1984) Analysis of multilocus pedigree data by computer. Histocompatibility testing, Albert ED, Baur MP, Mayr WR (eds), Springer-Verlag, Heidelberg, pp 52-58
Ohayon E, Cambon-Thomsen A (1986) Génétique des populations humaines. Colloque INSERM vol. 142, INSERM (ed)
Ohayon E, Cambon-Thomsen A (to be published) Genetic markers in French provinces, INSERM-DOIN (eds), Paris
Piazza A, Menozzi P (1983) Geographic variation in human gene frequencies. In: Numerical taxonomy, NATO ASI series, Springer-Verlag, Berlin, vol. G1, pp 444-450
Piazza A (1986) The genetic data from the French provinces : a tentative summary. In: Génétique des populations humaines, Ohayon E, Cambon-Thomsen A, (Eds), Colloque INSERM vol. 142, pp 345-350
Prevost P, Busson M, Marcelli-Barge A, 13 France-Transplant laboratories (1984) Distribution of HLA-A, B alleles in 13 panels of blood donors in France. Tissue Antigens 23 : 301-307
Ryder LP, Andersen E, Svejgaard A (1978) An HLA map of Europe. Hum. Hered. 28 : 171-200
Salmon Ch, Cartron JP, Rouger Ph (1984) The human blood groups. Masson publishing USA (ed)
Vergnes H, Constans J, Quilici JC, Lefèvre-Witier Ph, Sevin J, Stevens M (1980) Study of red blood cell and serum enzymes in five Pyrenean communities and in a Basque population sample. Hum. Hered. 30 : 171-180
Youinou P, Gueguen A, Halle L, Jouquan J, Salmon D, Le Menn G, Ruffie J, Salmon Ch (1983) Distribution of HLA antigens in two Breton communities. Tissue Antigens 22 : 348-353

Aknowledgments : This work was partly supported by grants from INSERM (CCP 81.8008) and DFG (Ba 660/5-2, Al 92/13-2). The authors wish to thank the blood banks and the local organizations whose help made this survey possible as well as all the families who kindly accepted to participate. They are also grateful to the CICT, Toulouse for help in computer data management. The secretarial assistance of D. Sicre and A. Gourdin is gratefully aknowledged.

Table 1. Quality of HLA-antigen definition

Excellent	Good	Rather Good	Poor
A1	A28	A30	A31
A2	A32	Aw33	Aw66
A3		Aw34	
A29	Cw1		Bw41
A11	Cw2	Cw6	Bw58
A23	Cw3	Cw7	Bw61
A24	Cw4		
A25	Cw5	B38	DRw8
A26		B45	DRw9
	B18	Bw47	DRw10
B7	B35	Bw52	DRw12
B8	B37	Bw53	DRw14
B13	B39	Bw60	DQw2
B14	Bw50		
B27	B51	DR1	
B44	Bw55	DR3	
B49	Bw56	DRw13	
Bw62	Bw57	DQw1	
Bw63	Bw4	DQw3	
	Bw6		
DR7			
	DR2		
	DR4		
	DRw11		
	DRw52		
	DRw53		

The quality of antigen definition was assessed according to the correlation coefficient (r) between computer and laboratory antigen assignment, the quality score (Q score) the normalized Q score (N-Q score) and the number of sera used (Sierp et al 1986).

Excellent : $r \geq 0.96$ and at least 2 excellent sera
Good : $r \geq 0.91$ and at least 3 good or rather good sera
Rather good : $r \geq 0.75$ and at least 2 good or rather good sera
Poor : $r < 0.75$ and/or less than 2 rather good sera

Excellent sera are defined by a N-Q score ≥ 0.5 and Q score value ≥ 8,
Good sera are defined by a N-Q score > 0.5 and Q score value 4-8,
Rather good sera are defined by a N-Q score 0.25 to 0.49 and Q score value ≥ 4.

Table 2. Allele or haplotype frequencies (%) for non MHC systems in the total sample "PF"

Genetic system	Allele or haplotype	Frequency %
ABO	A	26.2
N = 2672	B	6.0
	0	67.7
Rh	DCe	42.7
N = 2665	DcE	13.1
	Dce	2.0
	DCE	0.4
	dce	40.2
	dCe	0.9
	dcE	0.6
	dCE	0.0
MNSs	M	52.2
N = 2633	N	47.8
	S	32.8
	s	67.1
	MS	23.8
	Ms	28.4
	NS	8.4
	Ns	39.4
P	P1	53.1
N = 2650	P2	46.9
Kell	K	4.2
N = 2667	k	95.8
Pi	M1	61.3
(α_1-anti	M2	15.3
trypsin)	M3	9.9
N = 2707	M4	1.4
	M5	0.6
	S	9.6
	Z	1.4
	Var	0.5
Gc	1F	15.6
(DBP)	1S	53.9
N = 2713	2	29.9
	Var	0.6
Hp	Hp1	38.4
N = 2687	Hp2	61.6
Tf	C1	75.7
N = 2706	C2	19.0
	C3	4.5
	C Var	0.1
	D	0.1
	B	0.6

Km N = 2720	Km(1) Km(3)	9.1 90.9
Gm N = 2718	1 2 3 4 5 6 7 8 9 10 11 14	68.2 19.8 8.7 0.9 0.8 0.9 0.1 0.1 0.2 0.1 0.1 0.1

N = number of unrelated individuals
Var = rare variant
See text for definition of the genetic systems

Code	Gm haplotype nomenclature	
1	$Gm^{4;\pm 23;5*}$	$Gm^{f;\pm n;b*}$
2	$Gm^{1,17;;21,28}$	$Gm^{za;;g^1g^5}$
3	$Gm^{1,2,17;;21,28}$	$Gm^{zax;;g^1g^5}$
4	$Gm^{;\pm 23;5*}$	$Gm^{;\pm n;b*}$
5	$Gm^{4;\pm 23;}$	$Gm^{f;\pm n;}$
6	$Gm^{1,17;;5*}$	$Gm^{za;;b*}$
7	$Gm^{1,17;;5*,6}$	$Gm^{za;;b*c^3}$
8	$Gm^{1,17;;10,11,13,15,16}$	$Gm^{za;;b5b^0b3st}$
9	$Gm^{17;;21,28}$	$Gm^{z;;g^1g^5}$
10	$Gm^{1,2,17;;5*}$	$Gm^{zax;;b*}$
11	$Gm^{4;\pm 23;5*28}$	$Gm^{f;\pm n;b*g^5}$
12	$Gm^{1,17;;10,11,13,15,16,6}$	$Gm^{za;;b5b^0b3stc^3}$
13	$Gm^{4;\pm 23;21,28}$	$Gm^{f;\pm n;g^1g^5}$
14	$Gm^{1,17;;5*28}$	$Gm^{za;;b*g^5}$

$5* = 5,10,11,13,14$ $b* = b^1,b5,b^0,b3,b^4$

Table 3 - HLA allele frequencies (%) estimated by different methods and BF, C4A, C4B, GLO allele frequencies in the total sample

	I	II	III
HLA-A	N=2733	n=5526	n=5451
A1	13.77	13.37	13.64
A2	28.17	27.32	28.31
A3	13.18	13.14	13.28
A9	12.03	12.11	12.27
A10	4.34	4.24	4.32
A11	5.78	5.75	5.80
A29	6.61	6.75	6.58
Aw19.2	7.16	6.89	6.73
A32	3.27	3.27	3.35
Aw33	1.09	1.08	1.22
A28	3.67	3.68	3.67
A blank	0.93	2.40	0.83
HLA-C	N=2719	n=5527	n=5439
Cw1	3.55	3.73	3.70
Cw2	5.08	5.23	4.93
Cw3	8.95	9.24	9.89
Cw4	12.29	12.95	12.64
Cw5	8.90	9.25	8.74
Cw6	7.52	7.54	8.52
Cw7	20.10	20.31	22.39
C blank	33.61	31.75	29.20
HLA-B	N=2733	n=5505	n=5451
B5	8.00	8.17	8.21
B7	10.45	10.20	10.44
B8	10.04	9.81	9.99
B12	17.29	17.16	17.59
B13	1.57	1.58	1.61
B14	3.90	3.87	3.88
B15	6.23	6.15	6.16
Bw16	3.90	3.92	3.90
B17	4.12	4.07	4.10
B18	5.90	6.06	5.93
B49	2.17	2.21	2.07
Bw50	1.23	1.17	1.23
Bw22	2.40	2.36	2.40
B27	3.56	3.58	3.59
B35	9.21	9.16	9.61
B37	1.28	1.27	1.28
B40	5.50	5.42	5.06
Bw41	0.50	0.49	0.64
Bw47	0.38	0.40	0.42
Bw53	0.87	0.86	0.75
B blank	1.55	2.09	1.22

HLA-DR	N=2672	n=5437	n=5361
DR1	10.16	10.29	10.28
DR2	13.66	13.35	13.61
DR3	16.17	15.99	16.16
DR4	13.60	13.41	13.87
DR5	14.48	14.20	13.42
DRw6	10.06	9.91	8.75
DR7	15.17	14.99	15.26
DRw8	1.40	1.38	2.53
DRw9	1.00	0.98	0.73
DRw10	0.43	0.42	0.39
DR blank	3.87	5.07	4.98

BF	n = 5355
S	72.84
F	22.84
F1	3.07
S0.7	1.16
Rare	0.09

C4A	n=5188
A1	0.25
A2	5.18
A3	70.47
A4	5.57
A5	0.17
A6	3.43
AQ0	13.39
Rare	0.60
Duplication	0.94

C4B	n=5188
B1	71.32
B2	7.88
B3	1.06
B4	0.33
B5	0.21
B6	0.89
B7	0.10
BQ0	16.39
Rare	0.79
Duplication	1.04

GLO	n=4787
GLO 1	46.84
GLO 2	53.13
rare	0.02

N = number of unrelated individuals
n = number of haplotypes used in the family analysis
I allele frequencies (f) calculated from antigen frequencies in unrelated individuals (F) as $f = 1-\sqrt{1-F}$ and based upon local antigen assignment
II allele frequencies estimated by FAP from the same data
III allele frequencies estimated by FAP using the corrected data set after computer serological analysis.
BF, C4A, C4B and GLO allele frequencies were also estimated by FAP

Table 4. Most frequent HLA haplotypes in the total sample "PF"

A1	Cw7	B8	C4AQ0	B1	BfS	DR3
A3	Cw7	B7	C4A3	B1	BfS	DR2
A29	Cw–	B44	C4A3	B1	BfF	DR7
A23	Cw4	B44	C4A3	B1	BfF	DR7
A2	Cw5	B44	C4A3	BQ0	BfS	DR4
A30	Cw5	B18	C4A3	BQ0	BfF1	DR3
A1	Cw6	B17	C4A6	B1	BfS	DR7
A3	Cw4	B35	C4A2,3	BQ0	BfF	DR1
A2	Cw3	Bw62	C4A4	B2	BfS	DR4
A2	Cw1	B51	C4A3	B1	BfS	DR2

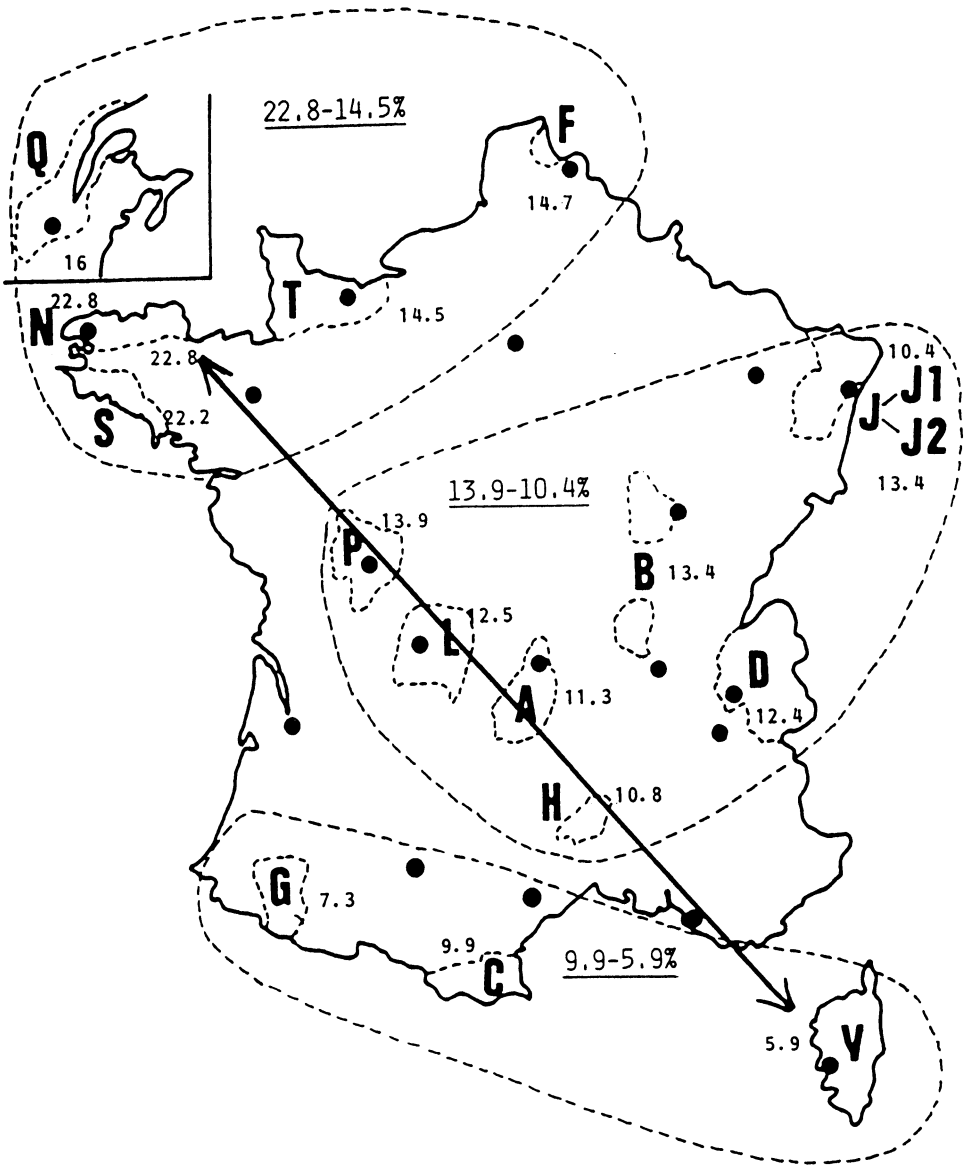

Fig 2. Distribution of HLA-DR4. Results are given as allele frequencies (%). For province code letter, see text.

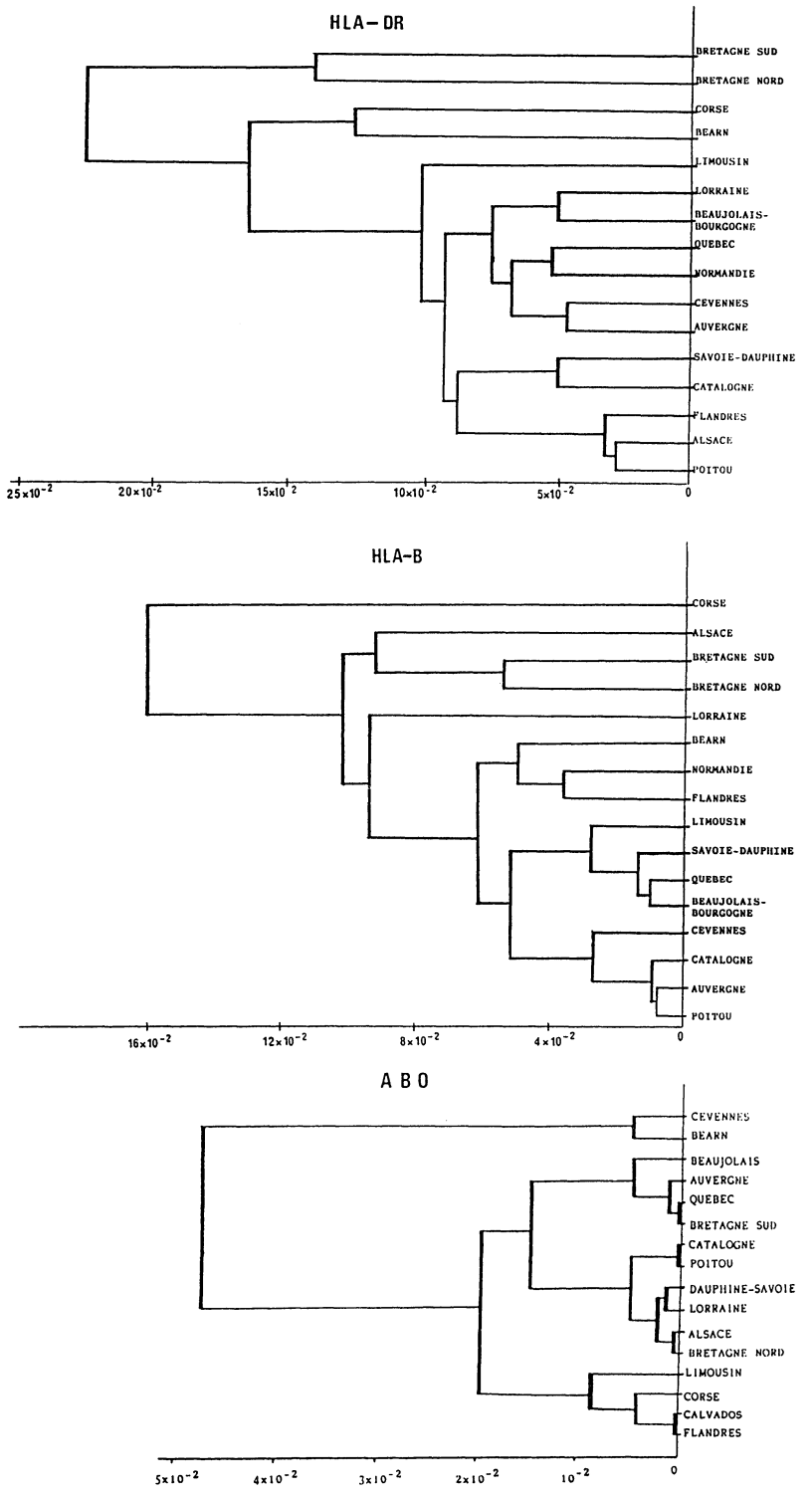

Fig 3. Dendrograms from X^2 genetic distances

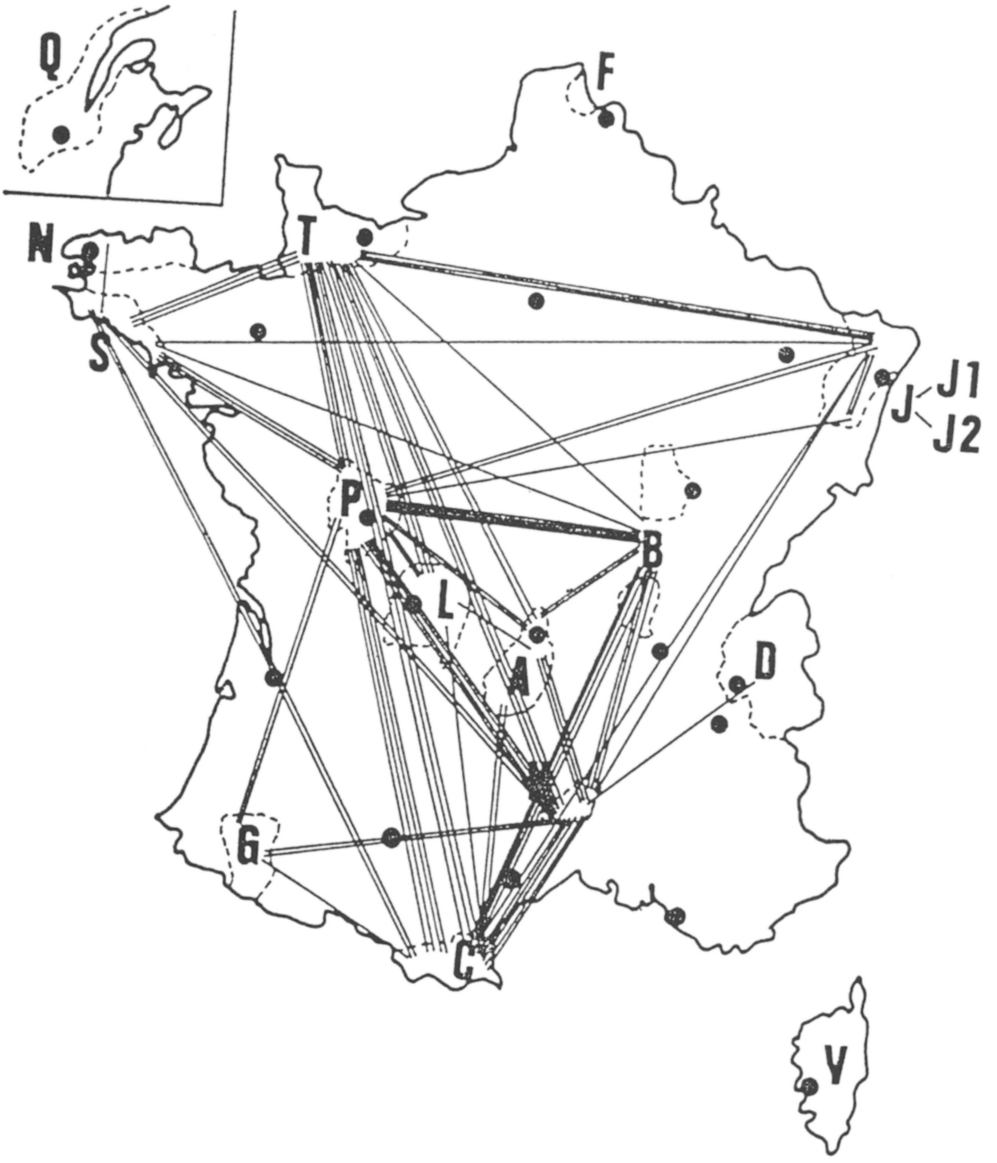

Fig 4. Genetic resemblances between French provinces.
The X^2 genetic distance was calculated for the following genetic systems : ABO, Rh, P, Kell, HLA-B, Gc, Pi, Tf
The number of lines between provinces is function of their genetic distances (1 : $0.07 < d < 0.08$; 2 : $0.06 < d < 0.07$; 3 : $0.05 < d < 0.06$; 4 : $0.04 < d < 0.05$; 5 : $0.03 < d < 0.04$). The closest regions are C (Catalogne) and H (Cévennes). For province code letter, see text.