ANNALS OF HUMAN BIOLOGY, 1983, VOL. 10, NO. 6, 505–522

Moors and Christians: an example of multivariate analysis applied to human blood-groups

R.A. REYMENT

Paleontological Institute, Uppsala University, Sweden

Received 27 October 1982; revised 18 May 1983

Summary. Published data on the frequencies of the alleles of the ABO, MNS, and Rh systems for populations in the western Mediterranean region are analysed by the multivariate statistical methods of canonical variates, principal components, principal coordinates, correspondence analysis and discriminant functions. It it shown that there is a 'Moorish substrate' in the eastern and north-eastern parts of Spain and in southern Portugal. Serological effects, such as could derive from the assimilation of a large Jewish population, cannot be identified in the data available. The theory that most Hispano-Moslems and Spanish Jews were of indigenous origin is not gainsaid by the serological data available.

1. Introduction

The increasing availability of blood-group data has, of recent years, been suggested by historians and anthropologists to be a potential source of primary information for reconstructing the historical development of human populations and one that should be free of subjectivity. Some preliminary studies have been made with rather far-reaching results. In interpreting such analyses, it should be kept in mind that they must be paired with a thorough background of historical knowledge.

The topic considered in the present paper is one pregnant with interest. There has been a great deal written about the ethnic influence of Islam on Spanish and Portuguese customs and institutions but little or nothing of a concrete nature concerning the composition of the present-day population of the Iberian Peninsula. The analyses accounted for here have as their goal the assessment of the value of serological data as an aid to historical reconstructions through the medium of multivariate statistical methods. It is, however, necessary to point out that published data on blood polymorphisms for Spain are far from being comprehensive — for example, analyses other than ABO frequencies are not publicly available for the crucial Andalusian region.

2. Historical background

The turbulent and colourful history of the Iberian Peninsula has captivated historians for hundreds of years. One of the most fascinating chapters concerns the Moslem era, which lasted, in identifiable form, for 900 years, being terminated officially by the edict of 1609 for the expulsion of the Moriscos.

In this paper, we shall be seeking traces of the Islamic period as registered in the blood-group frequencies of the present-day inhabitants of Spain and Portugal. Such an analysis could be expected to unveil a high degree of serological homogeneity over the entire Iberian Peninsula if the opinions of the older generations of Spanish historians hold true. There has, in the past, been a pronounced tendency to synonymize Christianity and the Visigoths with the good, and Islam and the Moors with the bad (Levi-Provençal 1932, p. 32). The most ridiculous manifestation of this

manner of thinking came to light in the sixteenth century with the decree of *limpieza de sangre* (purity of blood), which was still in effect as late as 1860. Purity of blood (i.e. neither Jewish nor Moorish roots) was a prerequisite for all official posts and for service in the Church. The required status of purity could, naturally, only be judged on hearsay as there were no adequate records (remembering that the time-interval involved was 900 years in the case of Moorish ascendancy and some 1600 years for Jewish admixture over a period marked by unrest and population upheavals). In fact, casting doubt on the purity of someone's 'blood' was a favourite way of getting a rival removed from the running or, at the very least, compromised, or having an adversary discredited.

The new generation of historians has begun to arrive at a more thoughtful evaluation of the demographical constitution of modern Spain and the following overview is based on their work together with that of the French orientalist Levi-Provençal (1932, 1967). In this connection, we note that Castro (1965) has been particularly critical of the tendency of older workers to obscure the populational history of Spain. Medina (1980) has exposed this point as well, but also brought to light that the Islamic presentation of the history of Moslem Spain is challengeable.

Earliest North African and Semitic immigration

Officially, the Moorish era in Spain and Portugal begins with the landing of Tarik (whose name is commemorated in 'Gibraltar' = Jabal Tarik) in the Bay of Algeciras. However, as several historians have pointed out (e.g. Rodriguez 1979, Medina 1980), the southern and eastern shores of the Peninsula were subjected to frequent episodes of colonization by Semitic groups (Jews, Phoenicians and Syrians), as well as North Africans, both before and during the Roman period; the North African influx was greatest in the Betic region of Andalusia. During the Visigothic era (mainly the sixth and seventh centuries), there was free cultural exchange between Andalusia and North Africa (Medina 1980, p. 149). The 'Moors' were a union of exiled Andalusians (what today would be called 'freedom fighters') and North Africans.

As regards the origin of the Iberians, opinions are sharply divided. To Medina (1980), the Iberian element in Spain was no more than a cultural manifestation (see also Parellada 1980). To others (*Gran Enciclopedia de Andalucía* 1979–1981, p. 961), the Iberians were a definite people who are thought to have entered Spain from North Africa and who are assumed by some to have been related to the Berbers. The Visigothic element in Spain never exceeded 2% of the population; in Andalusia, it was virtually absent (*Gran Enciclopedia de Andalucía* 1979–1981, p. 3295).

The Moslem invasion

In the year 711, a small force of Arabs, Berbers and exiled Andalusians, under the captaincy of Tarik ibn Ziyad, crossed the Strait of Gibraltar and within a short period of time, and with the collaboration of a large segment of the native population, almost the entire extent of the Iberian Peninsula had fallen into their hands and the tottering unpopular regime of the descendants of the Visigothic minority had been overthrown. The army of invasion was, in fact, mainly non-Moslem. Tarik, the governor of Tangiers, which at that time was a Visigothic dependency, may himself have been of Visigothic origin (cf. Medina 1980, *Gran Enciclopedia de Andalucía* 1979–1981, p. 1461).

The second invasion force was led by the Arab Musa ibn Nusayr who extended

the area of conquest and consolidated the gains made by Tarik. One of the remarkable aspects of the invasion was the large-scale acceptance of Islam by the indigenous population. Medina (1980) interprets this as being the outcome of the similarity in early Moslem dogma and the beliefs of unitarian Christianity, adhered to by the vast majority of the inhabitants of the Peninsula and which was incompatible with the newly adopted Trinitarianism of the upper segment of Visigothic society (Medina 1980). Many converts (*Musàlima*, whose descendants were termed Muwallads) were, in fact, only partly assimilated and there is ample evidence to show that they united the superficial observance of Islamic rites with the practice of unitarian Christianity for more than a century (Castro 1965, p. 176, Medina 1980, p. 178, Sánchez 1981, pp. 80–83, Martín 1981, p. 104). Some of the Moslem immigrants embraced Christianity (the so-called *tornadizos*), not forgetting the traditionally Christian Arabs from the Middle East, some of whom were martyred in Moslem Spain. Other inhabitants of the Peninsula adopted both religions, according to whatever best suited their purposes (the so-called *enaciados*, who subsequently were to play an important rôle in the latter stage of the reconquest, and regional rememberance of whom still exists — Levi-Provençal 1967, pp. 183, 184, 217).

Whether Trinitarian Christians (Mozarabs, from 'Musta-ribun', the name for Trinitarians living under Moslem rule), Moslems, or enaciados, the descendants of the original inhabitants of the Iberian Peninsula tended to maintain a social distance from the people of immigrant origin, the Arabs, Syrians and Berbers (Castro 1965, p. 191). They spoke Romance, from which modern Spanish has evolved, although educated Muwallads and Mozarabs were usually proficient in Arabic, the language of learning. There were numerous armed conflicts and insurrections against the Arab ruling cadre and, in addition, Arabs and Berbers were frequently at odds with each other owing to the discriminatory practices of the former.

The population of the Iberian Peninsula at the time of the Moslem invasion is variously estimated as having been between two and four million people, of which the majority came to identify with Islam. During the first four centuries of Moslem rule, Mozarabs and Jews lived in harmony with the lords of the land. As the centuries passed, however, the power of the predominantly Christian Basque-dominated states of the far north increased (cf. Allières 1977, p. 23). Consequently, despite the importation of first the Almovarids in the eleventh century, and the Almohads in the twelfth century, religious fanatics from North Africa who introduced a hitherto unknown element of intolerance into daily life which sparked off the flight of most Mozarabs and Jews to northern Spain, the Kingdom of Granada, the last bastion of free Islam in the Iberian Peninsula, capitulated in 1492. At this time, only Granada and parts of Murcia and Valencia retained larger Arabic-speaking nuclei, the remaining Moors either having been assimilated imperceptibly over the centuries of Christian rule, or degenerated into nominal Moslems (*Mudéjares*), without knowledge of Arabic and only a rather vague idea of Islam (Lea 1901, pp. 55, 59, 151, Watt 1965, pp. 54, 56, García Arenal 1975, p. 103). Zoido (1981, pp. 318–325) has reviewed the assimilation process with particular reference to the Sevillian region and the lack of concern of officialdom with the faceless masses: only the leading personalities of conquered Moslem communities were deported. See also Carol (1976, pp. 237–238).

In an effort to rid Spain of the last vestiges of Islam, the Moriscos, as the forcibly Christianized (in 1502) survivors of the Kingdom of Granada were

known, were expelled over the years 1609 to 1614, the total number of deportees being claimed to stop just short of 300 000 persons, according to modern estimates, this being somewhat less than 10 to 20% of the estimated Moslem population of the eleventh and twelfth centuries (Lea 1901, p. 360, Garcia de Cortázar 1976, p. 75).

The expulsion exercise was only partly successful, first, owing to its monstrous arbitrariness and, secondly, because of the calculated subversion of the edict by landowners, who needed a plentiful supply of cheap labour for their estates; this was particularly so in the eastern regions where, as the zeal of the Holy Inquisition increased, so did the traffic in certificates of good Christian standing. Many Moriscos who were little interested in the active practice of Islam returned clandestinely to their native land, rather than face the uncertainty of life in North Africa. There were no physical differences between 'New' and 'Old' Christians, nor could the majority be identified from speech habits. A conspiracy of silence developed spontaneously from fear, and the step to the charnel houses of the Church was short for the garrulous (Cardaillac 1979, Caro 1976).

By 1618, the increasingly hopeless task of tracking down returned Moriscos, and the people who had kept themselves in hiding (*Gran Enciclopedia de Andalucía* 1979–1981, p. 3341), often masquerading as gypsies, was abandoned by the Church authorities (Garcia Arenal 1975) and a general sigh of relief was breathed, not the least by many an old Christian, as not a few of them had been enmeshed in the frenzy of deportation, the victims of unjust denunciations by jealous neighbours (Lea 1901, pp. 353, 357, 364, 365, 374).

Intermarriage between religious casts was not uncommon until the close of the eleventh century (Castro 1965, pp. 179, 209, Lea 1901, p. 344), but became rarer under the Almoravid occupation. During the 120 years of the Moriscan era in Spanish history, intermarriage was facilitated by the imposed religious homogeneity (Lea 1901, pp. 19, 21, 26, 344, Levi-Provençal 1932, p. 35, Cardaillac 1979, p. 48).

Exemptions from the edict of expulsion were granted to former members of the Granadan nobility who had become incorporated with the Christian aristocracy, members of the clergy, responsible members of society, the offspring of mixed marriages, and orphans (Lea 1901, p. 321, Cardaillac 1979, Dominguez 1979, pp. 142, 144–146, *Gran Enciclopedia de Andalucía* 1979–1981, Caro 1976, p. 226).

Portuguese Moriscos were but rarely subjected to expulsion from their country and then only because of a direct refusal in the fifteenth century to adhere to the Christian religion; this action was instigated in acquiescence to pressure from the Spanish Inquisition (Bourdon 1970, Livermore 1976). Thus, southern Portugal can be considered as a standard of reference for a largely intact Moorish population (a caveat here though: serological data for Portugal must be used with caution owing to the sizeable assimilation of black Africans over the past 400 years).

It is of interest to note that Moslem customs survive to this day in many parts of rural Andalusia (*Gran Enciclopedia de Andalucía* 1979–1981, pp. 210, 2309, 2353, 2454, 2768, 2780, 2809, 3230) and as late as 1769, a Granadan was punished for practising Islam (Lea 1901, pp. 392–393). See also Caro (1976, p. 246).

The Taifas

Figure 1 displays the regional subdivisions of the Iberian Peninsula into taifas (*tawáif*) after the collapse of the central power structure of the Cordoban caliphate. It will be seen that the eastern taifas were ruled by so-called 'slave dynasties', that is, people of mainly Christian origin (the *saqaliba*) and consisting of a fusion of the

elite military corps of Frankish, Negro, Slavic and Catalonian slaves, and other demographic imports (Ladero 1979, p. 101). Just how far down into society this mixture extended is not known.

There is a large central field in figure 1 which was in the hands of Muwallads, and several smaller ones to the south which were ruled by dynasties of Arab, Berber or Syrian origin. It will be shown that these regional distributions seem to be of significance for the statistical discussions. It should also be mentioned that there was an appreciable import of black African household slaves and concubines into Moslem areas (Levi-Provençal 1932, p. 31, Watt 1965, p. 91, *Gran Enciclopedia de Andalucía* 1979–1981, p. 308, Caro 1976, pp. 87–91).

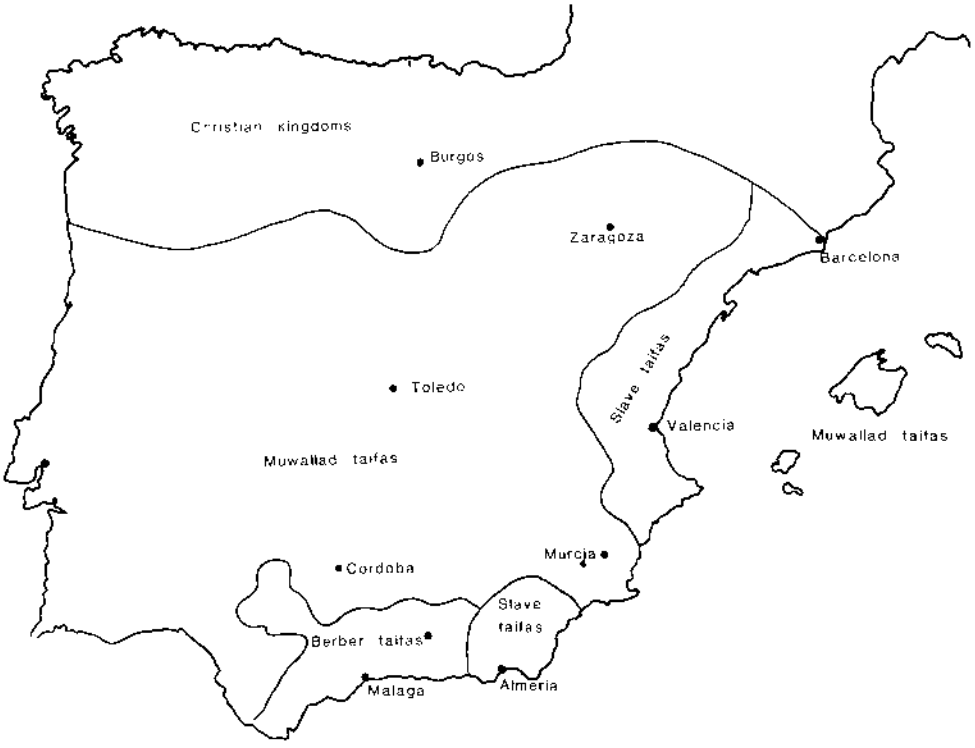


Figure 1. Sketch map showing the approximate distribution of the main taifas around the eleventh century (after Sánchez 1981, p. 128).

Spanish and Portuguese Jews

Spain had, at the time of the invasion from North Africa in 711, a well-established relatively large and successful Jewish population (Torroba 1967, Levi-Provençal 1967, p. 230, Bourdon 1970, p. 44, Domínguez 1977, Porcel 1977, Medina 1980, Baer 1981, Valdeón 1981, p. 28). Medina (1980, pp. 441–459) argues that most of the Iberian Jews descended from proselytes. Those Jews who refused to adopt Christianity after the fall of Granada in 1492 were expelled. Baptism was chosen by many and these unwilling converts swelled the vast group of *conversos* from the pogroms and mass persecutions of earlier years. At first, Spanish Jews could find refuge in Portugal, where, however, they, too, soon fell victims to the zeal of the Holy Inquisition. Some fled to countries round the Mediterranean (where

there are still Spanish-speaking Jewish communities), others to the Netherlands and Britain, for example, the Disraelis (Torroba 1967, Mourant, Kopeć and Domaniewska-Sobczak 1978, Baer 1981).

Analytical implications of Spanish history

The historical background of Spain suggests the following serologically based hypotheses.

(1) The populations of the eastern coast of Spain can be expected to deviate from most other parts of the Iberian Peninsula with respect to the frequencies of B and cDe (reflecting traces of the slave-taifas, superimposed on earlier colonization episodes ((Parellada 1980)) — the pre-Moslem history of the region has a record of frequent Semitic incursions).
(2) The population of southern Portugal should display identifiable 'Moorish characteristics' if, indeed, Moors originally possessed typical frequencies of blood polymorphisms.
(3) The wide distribution of descendants of the *conversos* should bear some relationships to blood-group frequencies of modern Sephardic communities, unless most Iberian Jews descended from proselytes, as thought by Medina (1980, p. 441).

3. The data and methods

All data derive from the tables published in the standard references of Mourant, Kopeć and Domaniewska-Sobczak (1976, 1978). For obvious reasons, the material available varies in geographical coverage and statistical completeness (cf. note by Piazza, Menozzi and Cavalli-Sforza 1981). The ABO system is the best represented, as to be expected, whereas the rhesus polymorphisms are much less extensively documented. The lack of Rh determinations for Andalusia is an example.

Multivariate statistical methods have not been used to any extent in the study of human serological data, although many interesting situations are highly multivariate. Recent applications have, however, been made by Carmelli and Cavalli-Sforza (1979) for Jews and a multivariate methodological problem has been considered by Piazza *et al.* (1981). Triangular diagrams showing ABO frequencies are not multivariate statistical in the strict sense and, besides, they suffer from the problem of closure in that the frequencies must perforce sum to a constant (cf. Mourant *et al.* 1978); as has been shown by Aitchison (1982), such plots may distort relationships between sample points.

The great advantage of multivariate statistical analysis is that it marshals the small differences in the observations which, when taken collectively, may be significant, but when viewed each on its own, may not be diagnostic. In the special case of frequency data, such as blood polymorphisms, there is a need for particular care. These polymorphisms represent discrete variables in the sense that part of an allele cannot be inherited. Consequently, if a population has zero representation of some allele, say B for Basques, it would be misleading to embark upon a multivariate analysis of a set of polymorphisms, as the ordinating effect of multivariate techniques would tend to blur the diagnostic significance implied by $B = 0$.

Some useful technical references for the multivariate methods used here are Blackith and Reyment (1971), Benzécri (1973), Jöreskog, Klovan and Reyment (1976), Gnanadesikan (1977), Everitt (1978), Pimentel (1979) and Hawkins (1980).

Multivariate methods in serological analysis

In order to assure the successful application of multivariate methods to serological data, we need to proceed by means of an analytical strategy. As already noted, a prime concern is that of the constant-sum condition of measurements on blood polymorphisms. For example, if one has determined A and B for some population, the value for O can be obtained by simple subtraction. Such variables occur in a closed simplex space for which standard methods of multivariate statistical analysis are not immediately applicable. Aitchison (1982, 1983) has given the problem careful attention and suggested a way of transforming from closed simplex space to open Cartesian space. He advocates the use of the log-ratio transformation and this approach has been used in the calculations of the present paper where suites of frequencies are involved and in which curvature seems to occur (Aitchison 1983). (N.B. It is not possible to escape from the closure straitjacket by merely dropping one of the variables of a set.)

Multivariate statistical methods can be conveniently thought of as falling into three major categories:

(1) The analysis of presumably homogeneous data-sets.
(2) The analysis of relationships between sub-vectors of a presumably homogeneous population.
(3) Methods devised for comparing two or more populations.

Most of the analyses of the present paper use the first of these categories. The usual methods for studying such data are various versions of 'eigen-analysis', to wit, principal components (and surrogates) for analyses concerned with examining relationships between variables (R-mode) and some form of 'inverted' principal components, such as principal coordinates, for relationships between individuals of a sample (Q-mode). The properties of the various methods of interest are presented in Jöreskog *et al.* (1976).

An applied method which has found favour of recent years is that currently known as correspondence analysis (Benzécri 1973, Jöreskog *et al.* 1976). This procedure combines Q-mode and R-mode plots on the same figure, suitably scaled, and its use is almost exclusively graphical. A useful feature of this method is that it provides a visual aid for locating distances between points in relation to specific principal-component loadings of significance for the segregation of individuals, or groupings of individuals.

In the present connection, special attention has been paid to questions of robust estimation in multivariate analysis and the interpretation of atypical observations (cf. Piazza *et al.* 1981, Campbell and Reyment 1980).

Selection of data

The statistical analysis is, in the first instance, directed towards elucidating multivariate serological relationships in the Iberian Peninsula and North Africa. As far as possible, homogeneous samples of mean frequencies for the alleles utilized have been employed. In a few cases, the ABO frequencies from one published source have been combined with frequencies for polymorphisms from a different source, but obtained from the same population. Data from areas lying somewhat outside the main region of interest have been included for purposes of comparison and interpretation for the reasons outlined below (cf. table 1).

Canary Islands. Although administered as a province of Spain, the indigenous population has an underlying African origin, thought to be akin to Berbers (the so-

called Guanches). Samples of descendants of the original inhabitants have therefore been included in the study.

Sicily. The dominant population element of Sicily is believed to be North African (cf. Parellada 1980). Sicily should therefore be serologically analogous to the populations of some areas of the Iberian Peninsula.

Malta. The main element of the population of Malta is Arab–North African (Maltese is a dialect of Arabic). Serologically, the Maltese should resemble Sicilian and North African populations.

4. The statistical analysis

The results presented in this section are so structured that problems of increasingly higher dimensionality are considered successively.

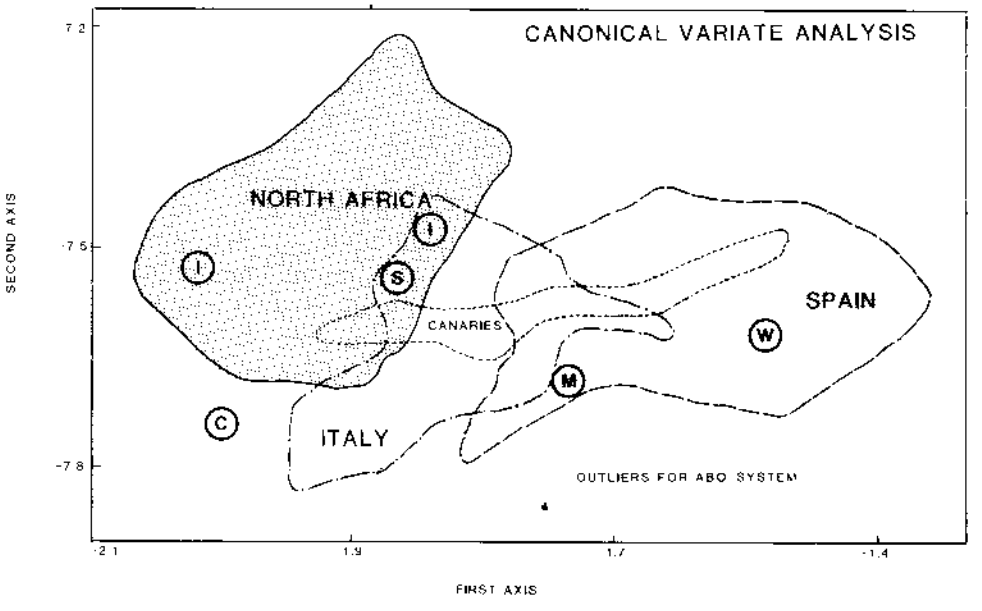


Figure 2. Projection of the sample observational vectors onto the plane of the first and second canonical variates for the ABO system. The key for the outlying values is: I, Italian (Sicily) samples; S, Spanish sample (Valencia); M, Morocco; W, Tunisia (Djerban Wahabites); C, Canary Islands.

The ABO system

The connections between populations of Spain, Algeria, the Canaries, Tunisia, Libya, Italy and Morocco were examined by canonical variate analysis in order to achieve a graphical impression of eventual overlap in the regions. Figure 2 shows the projections of the fields of the individual points onto the plane of the first and second canonical variates (see Blackith and Reyment 1971, for detailed case-histories of the application of canonical variate analysis in biology). Spain forms a group to the right encompassing one Tunisian mean (denoted W for Wahabites, an ascetic Moslem sect from Djerba, said to be pure Berber — cf. *Enzyklopaedie des Islams* 1913, p. 1175), some Canary Islander means, and a Moroccan mean. One Spanish mean (Valencia) lies among North Africans. The North African sets show considerable overlap and there is overlap between the Italian and North African fields (Sicilian means denoted I in figure 2).

Multivariate methods in serological analysis

In order to assure the successful application of multivariate methods to serological data, we need to proceed by means of an analytical strategy. As already noted, a prime concern is that of the constant-sum condition of measurements on blood polymorphisms. For example, if one has determined A and B for some population, the value for O can be obtained by simple subtraction. Such variables occur in a closed simplex space for which standard methods of multivariate statistical analysis are not immediately applicable. Aitchison (1982, 1983) has given the problem careful attention and suggested a way of transforming from closed simplex space to open Cartesian space. He advocates the use of the log-ratio transformation and this approach has been used in the calculations of the present paper where suites of frequencies are involved and in which curvature seems to occur (Aitchison 1983). (N.B. It is not possible to escape from the closure straitjacket by merely dropping one of the variables of a set.)

Multivariate statistical methods can be conveniently thought of as falling into three major categories:

(1) The analysis of presumably homogeneous data-sets.
(2) The analysis of relationships between sub-vectors of a presumably homogeneous population.
(3) Methods devised for comparing two or more populations.

Most of the analyses of the present paper use the first of these categories. The usual methods for studying such data are various versions of 'eigen-analysis', to wit, principal components (and surrogates) for analyses concerned with examining relationships between variables (R-mode) and some form of 'inverted' principal components, such as principal coordinates, for relationships between individuals of a sample (Q-mode). The properties of the various methods of interest are presented in Jöreskog *et al.* (1976).

An applied method which has found favour of recent years is that currently known as correspondence analysis (Benzécri 1973, Jöreskog *et al.* 1976). This procedure combines Q-mode and R-mode plots on the same figure, suitably scaled, and its use is almost exclusively graphical. A useful feature of this method is that it provides a visual aid for locating distances between points in relation to specific principal-component loadings of significance for the segregation of individuals, or groupings of individuals.

In the present connection, special attention has been paid to questions of robust estimation in multivariate analysis and the interpretation of atypical observations (cf. Piazza *et al.* 1981, Campbell and Reyment 1980).

Selection of data

The statistical analysis is, in the first instance, directed towards elucidating multivariate serological relationships in the Iberian Peninsula and North Africa. As far as possible, homogeneous samples of mean frequencies for the alleles utilized have been employed. In a few cases, the ABO frequencies from one published source have been combined with frequencies for polymorphisms from a different source, but obtained from the same population. Data from areas lying somewhat outside the main region of interest have been included for purposes of comparison and interpretation for the reasons outlined below (cf. table 1).

Canary Islands. Although administered as a province of Spain, the indigenous population has an underlying African origin, thought to be akin to Berbers (the so-

Frequencies of ABO, M and D

It is well known that the MNS allele differs but slightly on a regional scale (Beckman 1959, Mourant *et al.* 1976). There are not many data available compared with the ABO system and only 11 sets could be analysed in the present connection.

The results for the analysis of the multivariate means by principal coordinates are presented in figure 3. It will be seen that there seems to be a linear relationship between Ibiza. Djerban Wahabites and southern Portugal. Leon is located far from Asturias and Sevilla, both of which are located in the same general area as occupied by Ibiza, southern Portugal and Wahabites. A further interesting result is that the points for Sicily and Lyautey, N.W. Morocco (one of the places settled by Morisco refugees in the seventeenth century) lie close to each other. A remarkable feature of this projection is that the means for Sephardic Jews of Morocco and Tunisia (Djerba) plot so far from each other (cf. Mourant *et al.* 1978).

Inclusion of the rhesus cDe complex

The cDe complex of the Rh system is claimed to be diagnostic for elucidating relationships between western Mediterranean populations, as its occurrence in significantly elevated frequencies is considered to reflect black African elements. Figure 4 displays the chart of the first and second principal coordinate axes for 16 population means for A, B, O and cDe. This graph includes a Prim minimum spanning network, a device which acts as a useful aid for indicating nearness relationships in multidimensional space; it should be noted that such a tree is specific for a particular set of observations and should not be used for generalizations (other than with great caution). In essence, figure 4 confirms the results obtained for ABO frequencies, although certain interesting groupings seem to emerge. Thus, Aragon, Valencia and the Algarve (southernmost Portugal) connect to Malta and northern Tunisia (Wahabites). Sicily relates to Morocco and Malta and Moroccan Jews differ from the other Jewish samples considered, which, in turn, are similarly located.

A correspondence analysis was made for the combination ABO:D:cDe. As already observed, this method provides a means of displaying the interrelationships between the distribution of the individual population means and the variables specifying them — here, the five frequencies. The region for cDe attracts populations of Berbers, Valencians, Canary Islanders and Barcelonans. The area for high B loadings concentrates Jewish samples.

The scores for the third axis of the correspondence analysis of the ABO and cDe data are particularly informative. When superimposed on a map (figure 5), the scores for the eastern provinces of Spain are not far from those yielded by Berbers. The score for the Algarve (southern Portugal) lies within this range as do those for Sicily and the Canary Islands. Sephardic Jews sampled in Israel display the highest positive scores of all, whereas Moroccan Jews fall within the range for Berbers (Jewish values in figure 5 enclosed in brackets). The scores for north-western Spain and northern Portugal are similar and differ substantially from those obtained for eastern Spain, southern Portugal and North Africa. Of interest is the value for Lebanese Moslems of the Shi'ah sect as it is close to those for the north-western Iberian Peninsula.

ABO system and Rh complexes CDe cDE cDe cde

A more comprehensive analysis of the rhesus system supports the indications

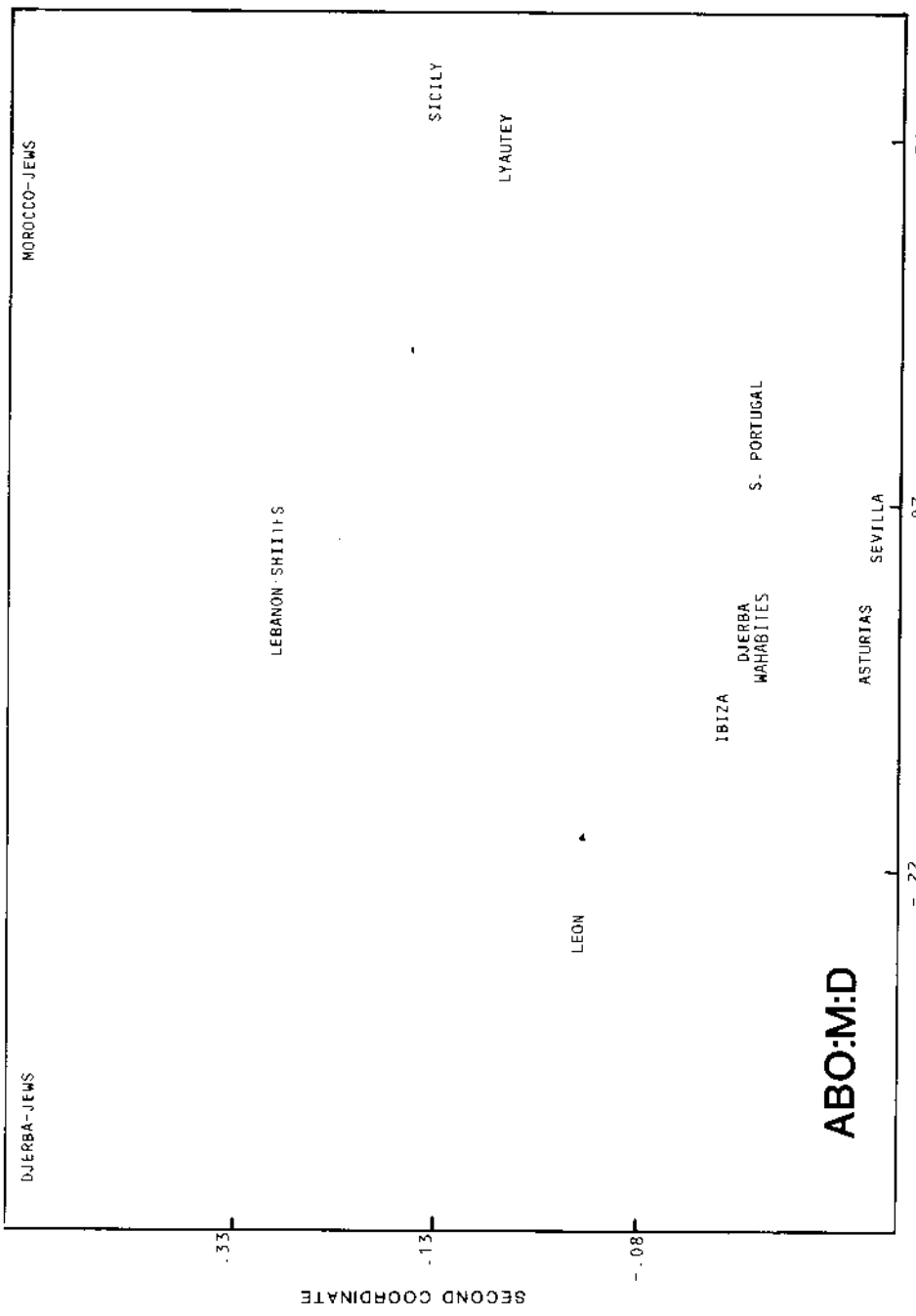


Figure 3. Principal coordinates chart for frequences of ABO, M and D.

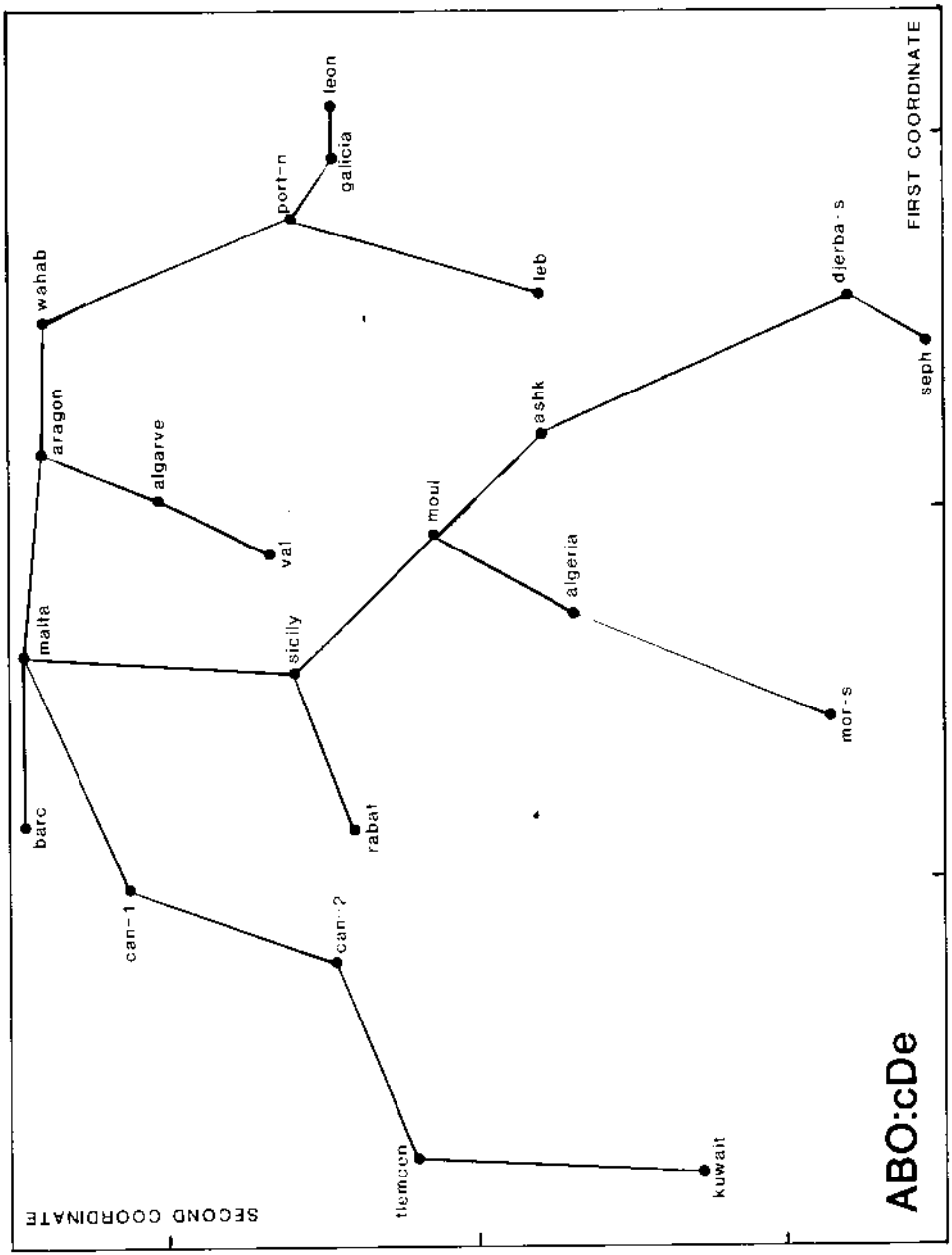
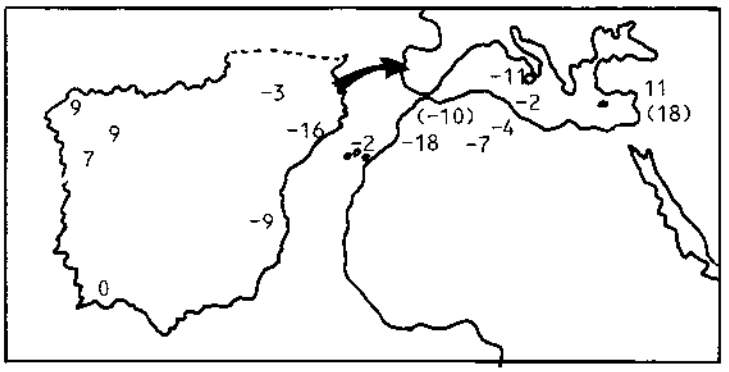


Figure 4. Principal coordinates graph for frequencies of ABO and cDE for 16 samples. Abbreviations: ashk, Ashkenazi Jews; barc, Barcelona; can, Canary Islands; djerba-s, Sephardic Jews from Djerba, Tunisia; gal, Galicia; leb, Shiitic Moslems from Lebanon; moul, Berbers from Moulay, Idriss; mor-s, Moroccan Jews; port-n, Northern Portugal; port-s, Southern Portugal; val, Valencia. Minimum spanning tree superimposed on plot.



THIRD AXIS OF CORRESPONDENCE ANALYSIS

$Z_3 = 0.15A - 0.02B - 0.07O + 0.13cDe$

Figure 5. Scores for individual populations of the correspondence analysis of ABO and cDe for the third axis shown in relation to geographical location. Values for Jews in brackets.

Table 1. Frequencies of blood polymorphisms used in the more highly multivariate analyses (extracted from Mourant *et al.* 1976)

Population	A	B	O	CDe	cDE	cDc	cde	M	D
Aragon (Spain)	27·30	6·49	66·21	44·07	10·48	3·18	37·11		
Ashkenazim	26·57	11·54	61·89			4·63			66·82
Asturias (Spain)	28·67	5·78	65·55					56·59	57·76
Barcelona (Spain)	20·94	3·97	75·09	41·85	12·71	3·30	40·52		
Basques — Spanish	24·17	1·30	74·53	40·32	4·31	7·26	43·62		
Catalonia	28·57	5·51	65·92	38·45	14·39	5·09	38·39		
Galicia (Spain)	32·18	6·71	61·11	43·22	12·72	3·97	38·27		
Galicia (Spain)	(30·57	6·21	63·22)	54·28	10·56	3·72	31·44		
Gran Canaria	25·16	7·95	66·89	45·18	7·71	9·73	32·87		
Gran Canaria	26·30	9·17	64·53	42·44	11·50	11·14	20·33		
Ibiza	29·97	4·99	65·04					50·17	64·21
Jews (Djerba)	29·72	15·16	55·12	45·21	10·60	4·88	35·19	69·29	64·49
Jews (Morocco)	19·96	19·11	60·93			5·77		55·91	65·74
Jews (Sephardic)	25·23	15·19	59·58	49·02	6·54	8·87	34·43		
Kuwait	15·33	15·47	69·00	38·12	15·39	14·42	20·92		72·67
Leon (Spain)	29·94	8·42	61·64	49·13	9·47	2·58	38·82		
Leon (Spain)	33·10	6·62	60·28	47·70	10·11	2·17	40·02	52·50	62·03
Lyautey (Morocco)	23·29	11·17	65·54					54·13	62·29
Malta	25·70	5·20	69·10	50·45	15·13	3·93	30·79		
Mouley Idriss (Morocco)	24·71	11·74	64·55	18·11	13·42	2·98	32·69		
Portugal (central)	29·32	6·47	64·21	43·45	11·90	4·76	37·87		
Portugal (north)	(28·95	5·84	65·21)	41·77	10·10	4·67	39·71	(52·34)	
Portugal (Oporto)	33·36	6·28	60·36	41·32	11·33	5·66	40·78		
Portugal (south)	26·93	6·44	66·63			5·66		52·09	63·50
Rabat (Morocco)	19·32	11·22	69·36			2·98			74·00
Sevilla (Spain)	27·61	6·70	65·69					55·97	54·90
Sh'iah (Lebanon)	29·09	10·82	60·09	50·36	11·48	7·55	28·43	53·58	69·88
Sicily	26·04	13·75	60·21	50·51	10·88	2·05	32·12		
Sicily	(23·46	9·70	66·84)	50·92	11·35	2·83	32·76		
Sicily	22·93	10·07	67·00			3·30		56·00	68·31
Tlemcen (Algeria)	18·23	9·90	71·87	56·15	6·97	3·58	33·30		
Valencia (Spain)	28·30	6·78	65·22	37·30	14·30	14·98	34·77		63·64
Wahabites (Djerba)	29·21	4·51	66·28			(2·00)		57·63	64·24

Bracketed values indicate analyses from different source but same area.

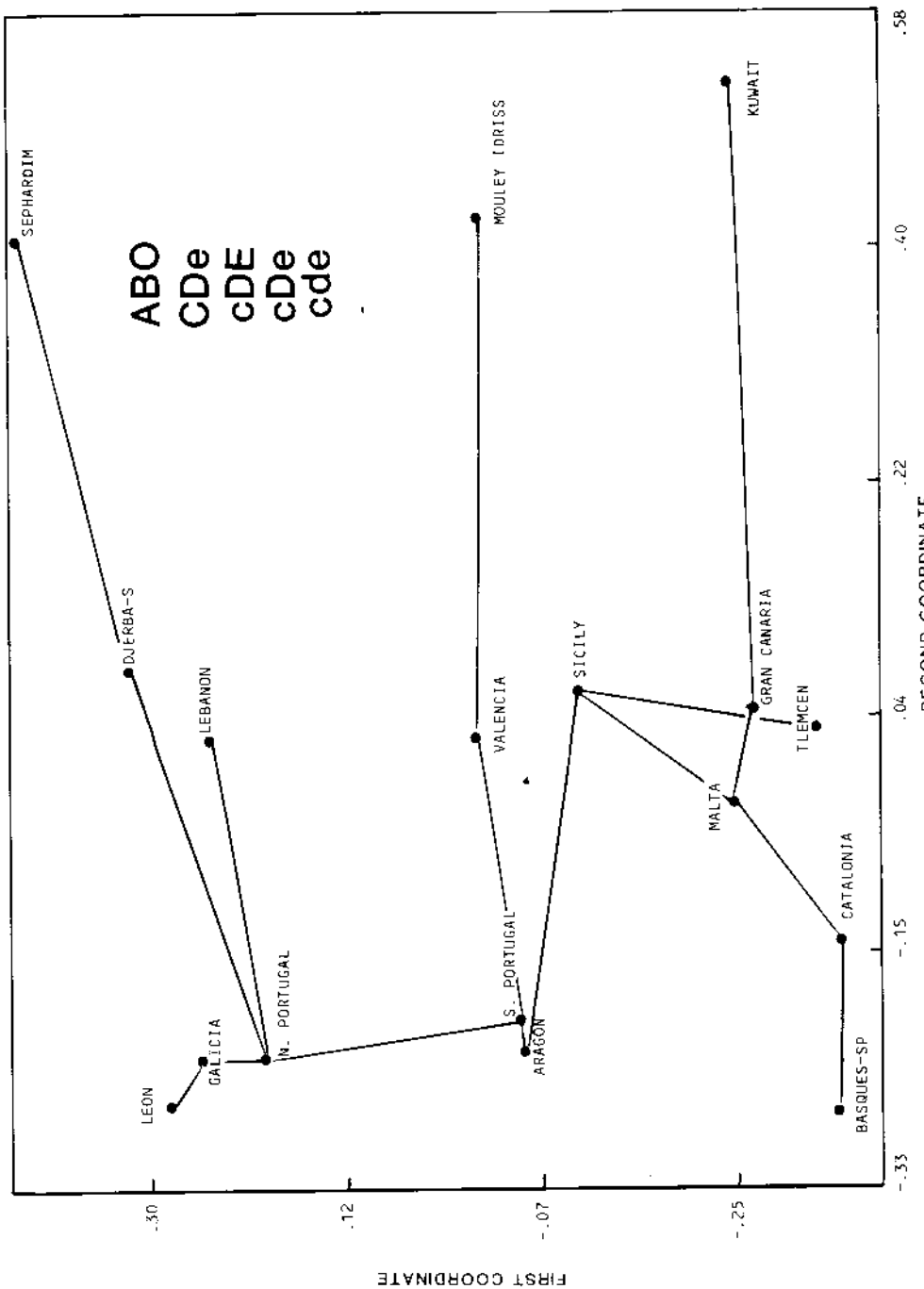


Figure 6. Plot of first two principal coordinates for ABO, CDe, cDE, cDe and cde. The branch of the minimum spanning tree between Sicily and Berbers (Tlemcen) passes, but does not cut, the connection between Malta and Gran Canaria.

already obtained. The plot of the first two principal coordinates for these variables (figure 6) tends to polarize further differentiation into (1) north-western Iberia, (2) eastern Spain and southern Portugal with (3) North Africa, Malta and Sicily, and Jews. Data for Basques were included here and it is perhaps significant that Catalonians and Spanish Basques do not lie very far from each other in this projection (cf. historical account in Parellada 1980). Could this imply that there is a proto-Basque element in Catalonia (Mourant *et al.* 1976, p. 64)?

The foregoing analysis was complemented by a principal-component study of 25 populations, including one of French Basques for which B = 0. In all essential details it supports the indications yielded by the coordinate analysis.

5. Interpretation of the analyses

Perhaps the most striking feature of the study is that the populations of Aragon and the eastern provinces of Spain deviate markedly from those of north-western Spain and northern Portugal, not only in regard to the ABO system, but also the Rh complexes, This is succinctly expressed by the map for the third vector of a correspondence analysis plotted against geographical location (figure 5).

The sample of Djerban Wahabites (Tunisia) is consistently allocated to the Spanish set in all analyses. Wahabites are a strict Moslem sect, founded by Muhammed ben Abd al-Wahhab in the eighteenth century. The Djerban Wahabites are believed to be pure Berbers (*Enzyklopaedie des Islám* 1913, p. 1175) and this, if true, could cast light on the populational history of Spain. I have been unable to establish any Morisco connection for the Djerban Wahabites.

It is well known that the majority of the originally relatively large Jewish population of the Iberian Peninsula was assimilated via mass conversions to Christianity. It is therefore surprising that this has not left identifiable traces in the blood polymorphisms analysed. Medina (1980, p. 441) considers most Spanish Jews to have descended from proselytes.

As far as can be judged from the published data, the frequencies of the blood polymorphisms analysed here for Galicia and Leon (north-western Spain) do not differ markedly from those occurring in northern Europe. The north-western part of the Iberian Peninsula was only sporadically occupied by the Moslems during the first 200 years of Al-Andalus.

6. Summarizing remarks

The multivariate statistical analysis of blood polymorphisms for the western Mediterranean indicates the existence of what might be termed a 'Moorish substrate' in some parts of the Iberian Peninsula. These traces are most marked in areas known to have harboured a relatively high proportion of exotic poulational elements, such as the slave taifas (largely as the result of elevated frequencies of B and cDe). On the information available, this condition is most strongly expressed in the eastern provinces of Spain and in southern Portugal. Given that the Moors of the Iberian Peninsula were racially little different from the Christians, it would seem that any identifiable Moorish traces must come from the North African, Middle East and black African components in the original Moslem population. The Moorish aspect of eastern Spain would have been reinforced by pre-Islamic immigration from North Africa and the Middle East (Medina 1980, Parellada 1980).

Castro (1965) has surmised that the earliest areas of Christian reconquest can be expected to display the highest degrees of assimilation of the original (Moslem)

populations. This theory is supported by the differences occurring between the populations of north-eastern Spain (early reconquest) and north-western Spain (only sporadically occupied by Moslem forces).

Can anything be said about the period at which most of the assimilation occurred? This would seem more likely to have taken place before the twelfth century, when religious tolerance was at its greatest, and before the hardening of the official Christian attitude brought about by the uprisings of the rapaciously exploited Mudéjars in southern Spain.

The Wahabite frequencies plot consistently with the samples from north-western Spain, which might indicate that 'pure Berbers' do not differ greatly from the populations of Leon and Galicia with respect to the polymorphisms considered.

Mediterranean areas with a similar past to that of the Iberian Peninsula behave similarly in the multivariate analyses (Malta and Sicily).

The methods of analysis used here lack, perforce, subtlety, owing to the restrictions imposed by the data available. Although the analyses cannot yield unequivocal solutions for the palaeodemography of the Iberian Peninsula, they do aid in pointing to new avenues of applied research such as, for example, the need for more comprehensive serological data for Andalusia. In some cases, the multivariate results could have been anticipated from mere inspection of B and cDe. The power of the multivariate approach becomes apparent first when there are no clear-cut dominances in the frequencies and in cases where the information required is contained in some natural ordering of the data.

Acknowledgements

I wish to thank Professor M. Hecht (New York) and Dr A. E. Mourant (Jersey), as well as many Spanish colleagues, among them in particular Professor J. Gonzalez Donoso (Málaga) and Dr M. Palacios Lamolda (Bilbao), for their interest and encouragement. I am indebted to Mr J. Aranki for aid in translating Arabic texts. The work was partly supported by Grant 2320–116 of the Swedish Natural Science Research Council (NFR).

References

AITCHISON, J., 1982, The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society*, **B44**, 139–177.

AITCHISON, J., 1983, Principal component analysis of compositional data. *Biometrika*, **70**, 57–65.

ALLIÈRES, J., 1977, *Les Basques* (Paris: Presses Universitaires de France).

BAER, Y., 1981, *Historia de los Judios en la España Cristiana* (Madrid: Altalena), 2 vols.

BECKMAN, L., 1959, A contribution to the physical anthropology and population genetics of Sweden. *Hereditas*, **45**, 1–189.

BENZÉCRI, J. P., 1973, *L'Analyse des Correspondances* (Paris: Dunod).

BLACKITH, R. E., and REYMENT, R. A., 1971, *Multivariate Morphometrics* (London: Academic Press).

BOURDON, A. A., 1970, *Histoire du Portugal* (Paris: Presses Universitaires de France).

CAMPBELL, N. A., and REYMENT, R. A., 1980, Robust multivariate procedures applied to the interpretation of atypical individuals of a Cretaceous foraminifer. *Cretaceous Research*, **1**, 207–221.

CARDAILLAC, L., 1979, *Moriscos y Cristianos: un Enfrentamiento Polémico (1492–1640)*. (Madrid: Fondo de Cultura Económica).

CARMELLI, D., and CAVALLI-SFORZA, L., 1979, The genetic origin of the Jews: a multivariate approach. *Human Biology*, **51**, 41–61.

CARO BARAJO, J., 1976, *Los Moriscos del Reino de Granada* (Madrid: Ediciones ISTMO).

CASTRO, A., 1965, *La Realidad Histórica de España* (México: Editorial Porrua SA).

DOMÍNGUEZ ORTIZ, A., 1979, Andalucia en el imperio español. In *Aproximación a la Historia de Andalucía*, edited by J. A. Lacomba, (Barcelona: Editorial Laia), pp. 131–155.

Estas conclusiones están de acuerdo con la hipótesis de que la mayor parte de los moros españoles eran muladíes, es decir, hispanos que se convirtieron al Islam (véase, por ejemplo, Medina (1980) y Castro (1965)).

A pesar de la asimilación de la mayoria de la población judía (los conversos), no aparecen efectos claros de esto en las frecuencias de los genes considerados, lo que tal vez pueda corroborar las conclusiones de Medina (1980): la inmensa mayoría de los judíos de la Península era indígena.

Las poblaciones del Mediterráneo occidental con un antepasado 'arabe' (Sicilia, Malta) se comportan en los análisis en manera análoga a las de la Península Ibérica.

مغاربة الأندلس ونصرانيوها: مثل على التحليل المتعدد المتغيرات مطبق على فصائل الدم البشري .

بقلم الدكتور ريتشارد ريّمنت استاذ قسم البليونتولوجية في جامعة ابسالا-السويد

الخلاصة

ان البيانات المنشورة عن تردد الجينة المورّثة بالدم بين انظمة العامل الريّصي ABO, MNS, Rh لسكان منطقة غرب البحر الأبيض المتوسط ، حللت بطريقة التحليل المتعدد المتغيرات بالقوانين المتنوعة. لقد تبيّن وجود "تحتية مغربية" في مناطق شرقي وشمال شرقي اسبانيا ، وفي جنوب البرتغال ومن المحتمل في غرب الأندلس .

ان تردد الجينة المورّثة بفصائل الدم الموجودة بعينات سكان شمال غرب اسبانيا ، وشمال البرتغال ، لا تختلف كثيرا عن هؤلاء من سكان اوروبا الشمالية .

تأثيرات الصول التي يمكن اشتقاقها من استيعاب عدد كبير من السكان اليهود (عن طريق الديانة بالمسيحية في القرنين الرابع والخامس عشر ميلادي) ، لا تعرّف من خلال البيانات المتوفرة .

النظرية التي تقول ان غالبية الأسبان ،المسلمين واليهود الأسبان ، كانوا من السكان الأصليين ، لا تتناقض مع الملاحظات المصولية .

DOMÍNGUEZ ORTIZ, A., 1977, *El Antiguo Regimen: los Reyes Catolicos y los Austrias.* Historia de España Alfaguera III (Madrid: Alianza Universidad Madrid).
ENZYKLOPAEDIE DES ISLAM, 1913, (Brill: Leiden).
EVERITT, B., 1978, *Graphical Techniques for Multivariate Data* (London: Heinemann).
GARCÍA ARENAL, M., 1975, *Los Moriscos.* (Madrid: Editora Nacional).
GARCÍA DE CORTÁZAR, J. A., 1976, *La Epoca Medieval.* Historia de España Alfaguera II (Madrid: Alianza Universidad Madrid).
Gran Enciclopedia de Andalucía, 1979–1981. (Sevilla: Promociones Culturales Andaluzas), 10 vols.
GNANADESIKAN, R., 1977, *Methods for Statistical Data Analysis of Multivariate Observations* (New York: Wiley and Sons).
HAWKINS, D. M., 1980, *The Identification of Outliers* (London: Chapman and Hall).
JÖRESKOG, K. G., KLOVAN, J. E., and REYMENT, R. A., 1976, *Geological Factor Analysis* (Amsterdam: Elsevier).
LADERO Q. M. A., 1979, La nueva Andalucía. In *Aproximación a la Historia de Andalucía*, edited by J. A. Lacomba, (Barcelona: Editorial Laia), pp. 101–130.
LEA, H. C., 1901, *The Moriscos of Spain* (London: Quaritch).
LEVI-PROVENÇAL, E., 1932, *L'Espagne Musulmane au X-ème Siècle.* (Paris: Larose).
LEVI-PROVENÇAL, E., 1967, *Histoire de l'Espagne Musulmane, III Le Siècle du Califat de Cordoue.* (Paris: Maisonneuve et Larose).
LIVERMORE, R. V., 1976, *A New History of Portugal* (Cambridge: Cambridge University Press).
MARTÍN, J. L., 1981, Reinos y condados del Norte. La alta edad media. Historia de España 3. *Historia*, **15**, (Special vol.): 89–129.
MEDINA MOLERA, A., 1980, *Historia de Andalucía: I, De los Origines al Califato Andalusi* (Sevilla: Biblioteca Ediciones Andaluzas).
MOURANT, A. E., KOPEĆ, A. C., DOMANIEWSKA-SOBCZAK, K., 1976, *The Distribution of Human Blood Groups and other Polymorphisms* (London: Oxford University Press).
MOURANT, A. E., KOPEĆ, A. C., DOMANIEWSKA-SOBCZAK, K., 1978, *The Genetics of the Jews* (Oxford: Clarendon Press).
PARELLADA DE CARDELLAC, J., 1980, *El Origen de los Vascos* (Barcelona: Plaza & Janés).
PIAZZA, A. MENOZZI, P., and CAVALLI-SFORZA, L., 1981, The making and testing of geographic gene-frequency maps. *Biometrics*, **37**, 635–659.
PIMENTEL, B., 1977, *Morphometrics* (Iowa: Kendall-Hunt).
PORCEL, B., 1977, *Los Chuetas Mallorquines* (Barcelona: Bruguera).
RODRÍGUEZ NEILA, J. F., 1979, En los origines de Andalucía: de Tartessos a la invasión árabe. In *Aproximación a la Historia de Andalucía*, edited by J. A. Lacomba (Barcelona: Editorial Laia). pp. 11.51.
SÁNCHEZ, M., 1981, Al-Andalus (711–1031). La Alta Edad Media. Historia de España 3. *Historia*, **15**, (Special vol.), 37–88.
TORROBA BERNALDO DE QUIROS, F., 1967, *Los Judios Españoles.* (Madrid: Private Edition).
VALDEÓN, J., 1981, La Baja Edad Media. Historia de España, 5. *Historia*, **17**, (Special vol.).
WATT, K. M., 1965, *A History of Islamic Spain.* Islamic Surveys, 4 (Edinburgh: Edinburgh University Press).
ZOIDO NARANJO, A., 1981, Conquista, genocido y colonización de Al-Andalus. In *Historia de Andalucía*, edited by A. Medina Molera, (Sevilla: Biblioteca Ediciones Andaluzas), pp. 235–285.

Address correspondence to: Dr R. A. Reyment, Paleontologiska Institutionen, Uppsala Universitet, Box 558, S 751 22 Uppsala, Sweden.

Resumen. Moros y Cristianos: aplicación del análisis multivariable a los grupos sanguíneous ABO, MNS y Rh de las poblaciones actuales de la Península Ibérica.

En este trabajo se han utilizado métodos de análisis multivariable (variables canónicas, componentes principales, coordenadas principales, análisis de las correspondencias y función discriminadora) para interpretar los datos publicados sobre las frecuencias de los alelos A, B, O, M, N y el sistems Rh en el Mediterráneo Occidental.

Se ha demostrado que existe un "sustrato moro" en las regiones orientales y nororientales de España, es decir, en las regiones de las taifas de esclavos (con frecuencias elevadad de B y cDe provenientes, probablemente, del elemento negro de estas taifas) y en lugares de la primera fase de la reconquista en el nordeste de España. Hay también indicaciones de un 'sustrato moro' en el Algarve. Además, parece que también las poblaciones de Andalucía occidental incluyen restos de antepasados 'moros', pero hacen falta más observaciones para confirmar esta hipótesis.

En lo que concierne la región del noroeste de España y el norte de Portugal, los resultados de los análisis indican que las poblaciones son, respecto a los genes analizados, muy parecidas a las europeas occidentales, incluso nórdicas.