TESTING HIERARCHICAL LEVELS OF POPULATION SUB-STRUCTURING: THE AZORES ISLANDS (PORTUGAL) AS A CASE STUDY

CRISTINA SANTOS*, AUGUSTO ABADE† AND MANUELA LIMA*

*Center of Research in Natural Resources (CIRN) and Department of Biology, University of the Azores, Ponta Delgada, Azores, Portugal and †Department of Anthropology, University of Coimbra, Coimbra, Portugal

Summary. The Azores archipelago (Portugal) is formed by nine islands whose relative positions define them as three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo). Using the father's surname of 187,398 individuals living on the nine Azorean Islands, a population analysis based on inter-island relationship and hierarchical organization was conducted. The relation between islands was investigated using summary statistics, analysis of molecular variance (AMOVA) as well as graphical methods. When the values of heteronymy were contrasted with values of gene diversity based on haplogroup frequencies of the Y chromosome, it was possible to verify that Graciosa and Sta. Maria appeared to have the least diverse populations, and that Flores, despite its smaller population size and geographical isolation, has considerably higher levels of diversity. As for inter-island relationships, the difficulty of directly interpreting summary statistics values was evidenced. The AMOVA revealed that only 0·77% of the variation in surnames can be attributed to among-island variation, a value that, although small, can be considered significant. Application of Malécot's model revealed that geographic distance has an important impact in the genetic structure of the archipelago. Monmonier's maximum-difference algorithm demonstrated that the most isolated island of the archipelago appears to be Graciosa, followed by the islands of the Western group and by Sta. Maria. After integrating values of summary statistics with results from AMOVA and graphical methods, a more accurate genetic profile of the Azores, highly supported by genetic data, has emerged.

Introduction

Simplicity in scientific theories is usually seen as a virtue and population genetics is no exception. Most discussions on the genetics of populations start with the simplest

description of a population as a very large, single collection of randomly mating individuals. From this simple description genetic attributes of populations should be deduced. However, many real populations do not fit this simple model (Hey & Machado, 2003). Real populations often have complex geographies and consist of many sub-populations that are connected by gene exchange or migration between them, which means that they are hardly ever genetically homogeneous (Hartl & Clark, 1997; Jorde, 1980). Moreover, since distinct populations have their own founders, history and dynamics, their present-day structure reflects a retrospective of the interaction of these multiple factors that have to be carefully considered in the interpretation of purely mathematical data.

An awareness of these complexities, and the goal of using evolutionary models that are as realistic as possible, has led many population geneticists to focus their efforts on what is usually called the 'structure' of natural populations. One of the strategies in the study of structured populations has been the traditional mathematical approach that relies on explicit models and summary statistics (Hey & Machado, 2003). However, results from summary statistics are usually difficult to interpret, since values depend on the population size, on the number of subdivisions as well as on the mutation rate of the specific genetic markers being used (Jorde, 1980; Flint *et al.*, 1999; Jorde *et al.*, 2000; Hey & Machado, 2003).

The Azores archipelago (Portugal), located in the Atlantic Ocean 1500 km from the European mainland, is formed by nine islands of volcanic origin. The relative position of these islands allows the definition of three geographical groups: Eastern (S. Miguel and Sta. Maria), Central (Terceira, Faial, Pico, Graciosa and S. Jorge) and Western (Flores and Corvo) (Fig. 1). With a total area of 2344 km^2, the Azores have presently a total population of 237,315 inhabitants, distributed in a very asymmetric way among the islands (INE, 2001). The settling of the archipelago, discovered and inhabited in the 15th century, was a slow process that lasted almost a century, starting with the islands of Sta. Maria and São Miguel, and later being completed by the islands of Flores and Corvo (Mendonça, 1996). According to historical records, the first settlers came mainly from various regions of mainland Portugal and the island of Madeira. However, there is clear evidence, both from historical (Tomaz, 1966–1969; Merelim, 1966; Mendonça, 1996; Gomes, 1997; Gregório, in press; Mesquita, in press) as well as molecular data (Santos *et al.*, 2003; Montiel *et al.*, 2005; Fernando *et al.*, 2005; Neto *et al.*, in press), that people of other European regions, individuals of Jewish origin and African and Moorish slaves also contributed to the peopling of the archipelago.

From the end of the 16th century until the third quarter of the 19th century, the population of the Azores underwent a significant increase. From then, and until about 1930, a slight reduction in the number of inhabitants was observed as a consequence of migratory movement that intensified during the 19th century. After 1930 there was a considerable increase in the number of inhabitants, which reached its maximum by 1960, after which the population started a marked diminution, once again as a result of migratory movement. This exodus of Azoreans during the second half of the 20th century resulted in some islands decreasing their population to less than a half in less than a century.
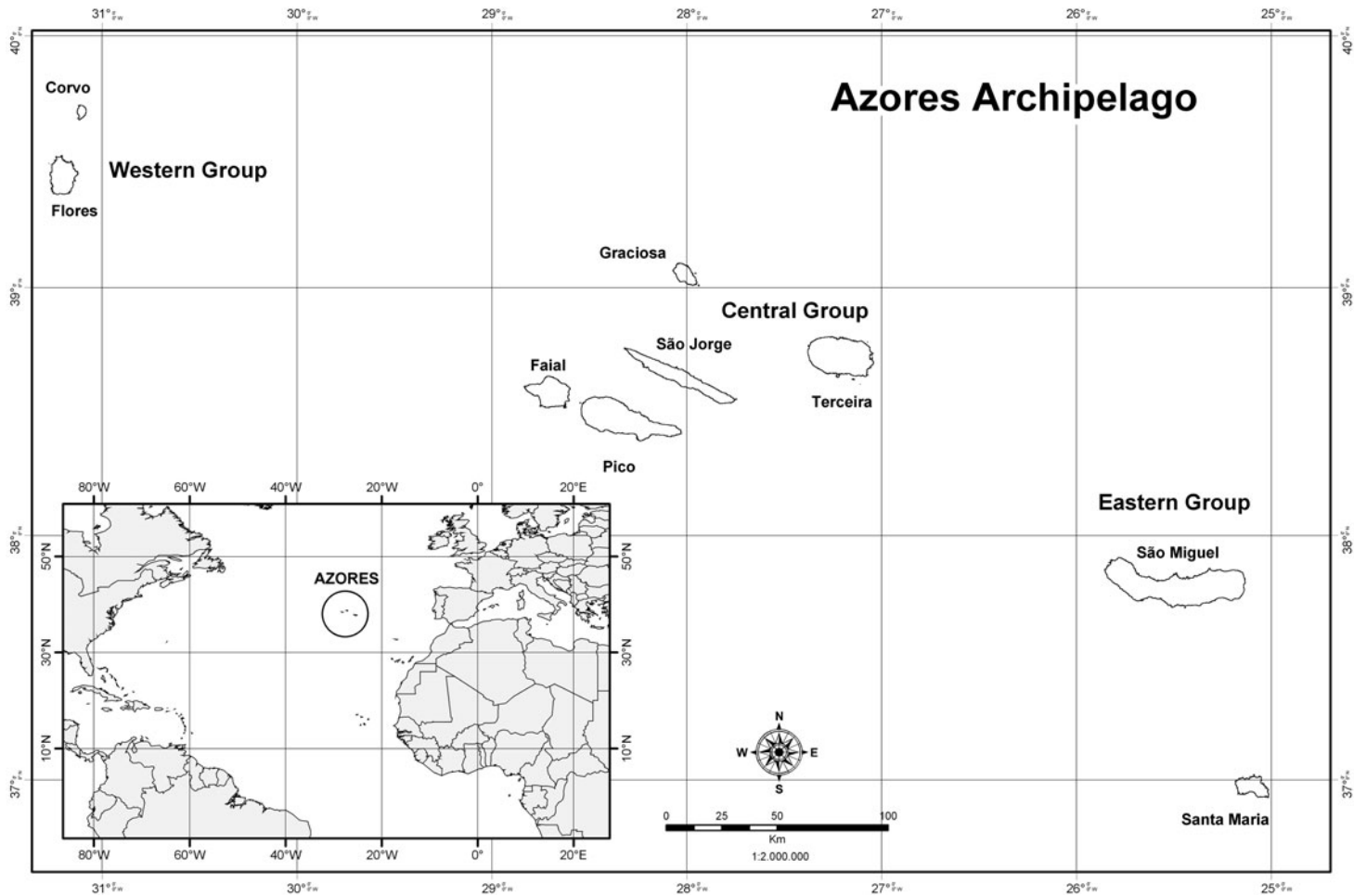
Fig. 1. Geographic location of the Azores. The archipelago is formed by three groups of islands: Western, Central and Eastern. Source: Geography Section, Department of Biology, University of the Azores.

Due to the complex and recent process of peopling, which has been affected by migration as well as its geographic characteristics, the Azores archipelago provides a suitable scenario to evaluate particular aspects of the impact of subdivision in human populations. The assessment of the population's sub-structure can be correlated with the existence of different founder populations for the distinct subdivisions, and can be used in the prevision of its theoretical consequences. Moreover, evaluation of the presence of a sub-structure and quantification of the admixture is an important prerequisite for selecting populations in which genes contributing to complex traits can be searched through association mapping (Ziv & Burchard, 2003; Freedman *et al.*, 2004; Marchini *et al.*, 2004; McKeigue, 2005). Detecting sub-structuring in populations is also particularly relevant in the context of forensic genetics, given its pertinence to the decision of introducing corrections in probability calculations and establishing the need for a specific forensic genetic database.

Research conducted so far, based on the monoparental (mitochondrial DNA and the non-recombining region of the Y chromosome) and autosomal genetic systems, has highlighted the fact that the archipelago cannot be considered a truly homogeneous population, since the existence of genetic sub-structuring between groups of islands has been demonstrated (Santos *et al.*, 2003; Montiel *et al.*, 2005; Spinola *et al.*, 2005). Furthermore, there is clear evidence of differentiation between islands within the same geographical group (Bettencourt *et al.*, 2006).

In order to test the hierarchical levels of sub-structuring in the population of the Azores, the analysis of a large number of autosomal markers using representative samples of each of the nine islands of the archipelago is required, a project that is currently being developed. Since surnames in many human populations are inherited through the paternal lineage, they can be considered and interpreted as a marker of the Y chromosome and, therefore, as a first approach, they can be used to provide a genetic profile of the Azorean islands. However, although surnames constitute an inexpensive marker to obtain the profile of genetic variation and genetic structure of populations (Relethford, 1988; Colantonio *et al.*, 2003), it is important to acknowledge that they provide only a rough view of the genetic scenario of populations, since polyphyletism and incorrect transmission of surnames can occur in variable frequencies.

In this work, using data on surnames, analyses of inter-island relationships and hierarchical organization of the Azorean populations were conducted. Results obtained were also used to discuss the adequacy of distinct mathematical models for the study of the sub-structuring of human populations.

Methods

Population and data

Surnames were extracted from the Azorean electoral census list for the year 2003. The father's surnames of each of the 187,398 individuals listed in the census were retrieved. The data used correspond to the surnames of the inhabitants of the Azores aged 18 years or older at the time of the census: 3243 for Flores, 339 for Corvo, 11,388 for Faial, 11,690 for Pico, 7943 for São Jorge, 3784 for Graciosa, 44,633 for

Terceira, 99,870 for São Miguel and 4508 for Sta. Maria. The surnames of all the individuals were pooled together in order to combine those which, although they had the same phonetics, were written differently.

Surnames and genetic structure

In order to evaluate the genetic structure of each island, heterozygosity – or heteronymy in the case of surnames (Harpending & Ward, 1982; Nei, 1987) – and Morton's intra-population *a priori* kinship ($Ø_{ii}$) were calculated (as defined in Relethford, 1988).

To assess the impact of population size on the intra-population kinship, the non-linear model of regression proposed by Dahlberg (1948), which allows the Morton intra-population *a priori* kinship to be related to the number of inhabitants (N_i), as defined below, was used:

$$Ø_{ii} = a\left(\frac{1}{N_i}\right)^b,$$

where *a* and *b* are the constants of the regression model.

Malécot's isolation-by-distance model (Malécot, 1948) was used to relate *a priori* kinship and geographic distance between islands. In terms of *a priori* kinship, the model states that kinship declines exponentially with distance, i.e.:

$$Ø_{ij} = ae^{-bd},$$

where the constant *a* is an estimation of the unweighted mean local kinship, and *b* is a measure of distance (*d*) decay.

The relation between islands was investigated using: (a) summary statistics (Wright, 1951; Relethford, 1988); (b) analysis of molecular variance (AMOVA), which is usually used in the analysis of data from genetic markers (Weir & Cockerham, 1984; Excoffier *et al.*, 1992; Weir, 1996); and (c) graphical methods.

The AMOVA procedure estimates genetic structure indices using information on the allelic content of haplotypes, as well as their frequencies (Excoffier *et al.*, 1992). The significance of the covariance components associated with the different possible levels of genetic structure (within individuals, within populations, within groups of populations, among groups) was tested using non-parametric permutation procedures (Excoffier *et al.*, 1992); in this paper 1000 permutations were used. As for graphical methods, metric multidimensional scaling (MDS) of genetic distance based on random isonymy, correspondence analysis (only surnames with an intra-island frequency equal to or higher than 2% were used) and Monmonier's maximum-difference algorithm (Monmonier, 1973) were used to display the relation between the islands of the archipelago. Multidimensional scaling and correspondence analysis display similarities and dissimilarities in a virtual space that corresponds to the plot itself. This approach is particularly suitable for identifying clusters or outliers that are informative about the sort of differentiation underlying the genetic variability. Nevertheless, when the analysis aims to provide a real graphic representation of genetic differences in geographic space, alternative approaches can be used. In this context, Monmonier's

Table 1. Frequencies of surnames that present an intra-island frequency equal to or higher than 5%

Surname	Flores	Corvo	Faial	Pico	S. Jorge	Graciosa	Terceira	S. Miguel	Sta. Maria
Bettencourt	0·12	0·00	1·17	2·51	5·15	8·83	0·85	0·19	0·40
Braga	0·46	0·00	0·07	0·04	0·01	0·00	0·06	0·34	5·01
Cunha	0·09	0·59	0·26	0·19	0·71	5·05	0·52	0·20	0·27
Fraga	1·20	5·60	0·38	0·44	0·03	0·03	0·10	0·03	0·02
Freitas	5·74	2·06	0·94	1·51	0·88	0·16	0·91	0·33	3·73
Medeiros	0·99	0·59	1·98	1·39	0·68	0·29	0·82	5·30	1·24
Melo	1·60	0·59	1·25	2·20	0·77	5·71	2·03	2·59	3·70
Nunes	1·26	5·60	0·68	1·09	1·71	0·11	1·09	0·20	0·51
Picanço	0·00	0·00	0·17	0·05	0·03	5·81	0·12	0·01	0·00
Pimentel	1·39	5·01	0·28	0·80	0·26	0·03	0·57	1·20	0·24
Silva	4·47	5·01	11·55	8·67	6·47	21·49	5·72	3·55	0·95
Silveira	2·13	0·00	3·07	3·06	6·07	0·42	1·21	0·13	0·18
Sousa	4·04	1·47	1·88	2·48	4·70	3·01	3·08	3·70	9·12
Other	76·50	73·45	76·33	75·58	72·54	49·08	82·93	82·23	74·62

maximum-difference algorithm (Monmonier, 1973) was implemented for the visualization, on a geographic map, of the trend data contained in matrices (in this paper the genetic distance based on random isonymy was used). The algorithm was applied to a geometric network that connected all the populations (sampled locations) using Delaunay triangulation (Brassel & Reif, 1973), finding the edges associated with the highest rate of change in a given distance measure.

ISOnymy3 (Abade, 2003), Arlequin 2·000 (Schneider *et al.*, 2000), Barrier (Manni *et al.*, 2004) and SPSS 12·0.1 (SPSS, 2003) were used for data analysis.

Results and Discussion

Surnames and estimations of diversity

Table 1 shows the thirteen surnames that present an intra-island frequency equal to or higher than 5%. In the islands of the Central group, *Silva* is the most frequent surname; in the remaining islands, the more frequent surnames differ. This result is in agreement with historical reports of distinct waves of settlers for the various groups of islands.

In Table 2, the number of different surnames, as well as values of heteronymy and intra-population *a priori* kinship ($\emptyset_{ii}$), are presented for each island. The islands of Graciosa and Sta. Maria present the lowest values of heteronymy and the highest intra-population *a priori* kinship. By contrast, the islands of Terceira, S. Miguel and Flores (present study and Santos *et al.*, 2005) exhibit the highest values of heteronomy and the lowest values of $\emptyset_{ii}$. The values of $\emptyset_{ii}$ are similar to those obtained by Branco & Mota-Vieira (2005) using surnames extracted from the telephone directory, representing a subset of only 57,387 individuals. When the values of heteronymy are

Table 2. For each island the number of inhabitants (INE, 2001), the number of individuals considered in the analysis (the father's surname of each individual was used), the number and percentage of surnames, values of diversity based on surnames or heteronymy, values of *a priori* kinship and gene diversity based on NRY haplogroup frequency are presented

Island	No. inhabitants	No. individuals considered	No. (%) surnames	Heteronymy	*A priori* kinship ($\emptyset_{ii}$)	Gene diversity NRY (N)[a]
Flores	3995	3243	298 (9·19)	0·9845	0·0038	0·8571 (21)
Corvo	425	339	87 (25·66)	0·9750	0·0055	—
Faial	15,063	11,388	570 (5·01)	0·9762	0·0059	0·8143 (21)
Pico	14,806	11,690	431 (3·69)	0·9795	0·0051	0·7333 (6)
S. Jorge	9674	7943	373 (4·70)	0·9760	0·0060	0·6818 (12)
Graciosa	4780	3784	210 (5·55)	0·9288	0·0177	0·3182 (12)
Terceira	55,833	44,633	1093 (2·45)	0·9874	0·0032	0·6471 (17)
S. Miguel	131,609	99,870	1433 (1·43)	0·9851	0·0037	0·5818 (64)
Sta. Maria	5578	4508	271 (6·01)	0·9712	0·0071	0·3202 (23)

[a]Data from Montiel *et al.* (2005) and Fernando *et al.* (2005).

contrasted with values of gene diversity (Nei, 1987), based on haplogroup frequencies of the Y chromosome (Montiel *et al.*, 2005; Fernando *et al.*, 2005) (Table 2), it is possible to verify that Graciosa and Sta. Maria also appear as the least diverse populations, and that Flores has considerably higher levels of diversity.

Since population size is one of the factors that can theoretically influence $\emptyset_{ii}$, and considering that population size is very different between islands, the non-linear model proposed by Dahlberg (1948), which describes the relation between kinship and population size, was applied. The majority of the islands present levels of $\emptyset_{ii}$ that are lower than those predictable by population size alone. The graphical representation of the relation again calls attention to the island of Graciosa, in which the value of intra-population kinship was much higher than would be predicted by population size alone (Fig. 2). This result indicates that the high kinship and reduced diversity of surnames observed in Graciosa cannot be explained by its small population size. However, this result has to be interpreted with caution, since the number of inhabitants explains only 6·4% of the variation observed in kinship. Taken globally, the previous result indicates that population size is not a crucial factor in shaping the genetic structure of the Azores islands.

Inter-island relationship and hierarchical analysis of sub-structuring

Considering the nine islands of the archipelago, the F_{ST} value (random component of inbreeding) obtained, based on *a priori* kinship (Relethford, 1988) is 0·0043, a value similar to that obtained using the isolation-by-distance model (Malécot, 1948) ($F_{ST}=0·0058$). Both values put in evidence, if considering the limits proposed by
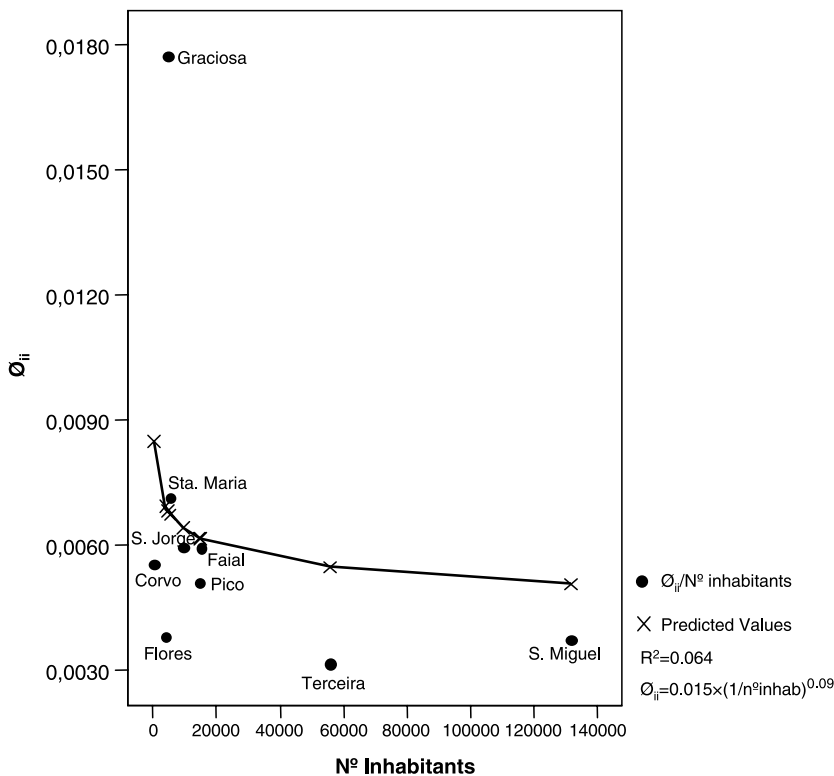
Fig. 2. Scatterplot of the number of inhabitants and the *a priori* kinship. The line of predicted values using the model of Dahlberg (1948) is shown.

Relethford (1988) based on surname analysis, the presence of medium levels of microdifferentiation. On the other hand, if the limits proposed by Wright (1984) are considered, the values of F_{ST} must be considered low. These seemingly contradictory inferences make obvious the difficulty in interpreting values of summary statistics. To obtain an insight into this problem, a molecular analysis of variance – AMOVA (Weir & Cockerham, 1984; Excoffier *et al.*, 1992; Weir, 1996) – was used to evaluate the significance of among-island variation.

Considering the nine islands of the archipelago, AMOVA results revealed that 99·23% of the variation in surnames corresponds to intra-island variation, whereas only 0·77% can be attributed to among-island variation, a value that, although small according to Wright, is nevertheless significant ($p<0.00001$). This indicates that at least one of the islands of the archipelago is set apart from the rest.

In Fig. 3 the multidimensional scaling (MDS) of genetic distance among islands, based on random isonymy, is presented. The clustering of populations can be related to the geography of the archipelago: the islands of the Western group are both located in the negative quadrant of dimensions 1 and 2; the islands of the Eastern group are both located in the negative quadrant of Dimension 1 and positive quadrant of Dimension 2; and finally, the islands of the Central group are located in
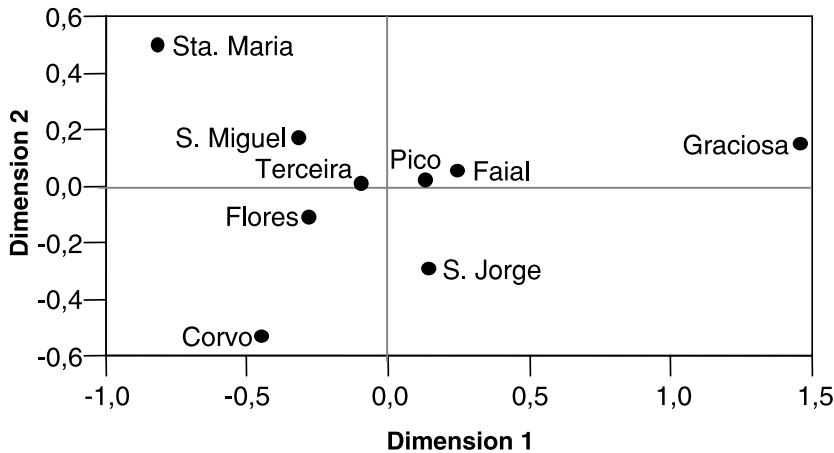
Fig. 3. Multidimensional scaling analysis plot representing the genetic distance based on a random isonymy matrix.

the positive quadrant of Dimension 1, with the exception of Terceira, which is located near the centre of the coordinates but in the negative quadrant of Dimension 1. Whereas Dimension 2 gives a distribution of islands mostly according to geographical groups and distances, Dimension 1 seems to reflect the peopling process. As is known, the peopling of the islands was initiated in the island of Sta. Maria, followed by the island of S. Miguel (both from the Eastern group), with individuals essentially from mainland Portugal. The islands of the Central group were peopled subsequently, it being well established that people of Flemish origin had an important impact in this process. Testifying this impact, HLA (human leukocyte antigen system) data (Spinola *et al.*, 2005) support the influence of Central European populations in the peopling of the Central group. This fact corroborates the positioning in the positive part of Dimension 1 of the islands belonging exclusively to this group of islands. The islands of the Western group were the last to be peopled, mainly with individuals that migrated from the other islands of the archipelago, this fact being in accordance with the position of these islands between the Eastern and Central groups.

Considering the islands individually (not taking into account the geographical grouping), four well individualized clusters can be identified (Fig. 3): one formed by the islands of S. Miguel, Flores, Terceira, Pico, Faial and S. Jorge and three other clusters constituted respectively by Sta. Maria, Graciosa and Corvo. Correspondence analysis results (Fig. 4) provide a similar picture with the islands of Sta. Maria and Graciosa being set apart. However, the islands of the Western group appear together in the plot with the island of the Central group. The differentiation of Sta. Maria and Graciosa, whose particular behaviour had been demonstrated in the MDS analysis, is justified by the presence of surnames on those islands that are almost absent on the other islands. In the case of Corvo (which in this analysis does not stand out from the cluster) the surnames present on this island appear to be essentially the same as those present on other islands.
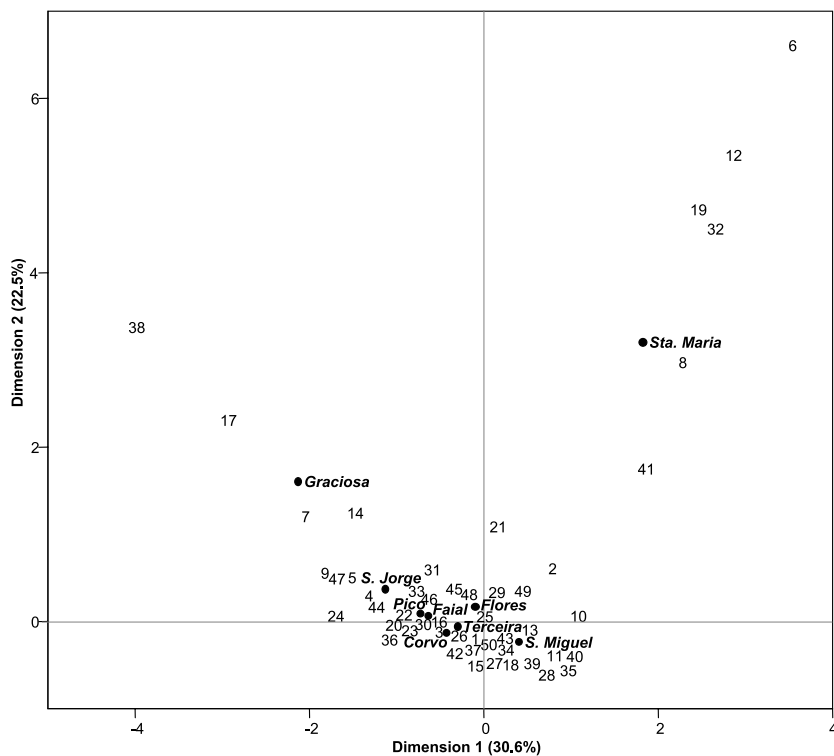
Fig. 4. Correspondence analysis, for islands and surnames (with a frequency equal to or higher than 2%). The first dimension accounts for 30·6% of the inertia, whereas the second accounts for 22·5%. 1, Alves; 2, Andrade; 3, Avelar; 4, Ávila; 5, Azevedo; 6, Bairos; 7, Bettencourt; 8, Braga; 9, Brasil; 10, Cabral; 11, Câmara; 12, Chaves; 13, Costa; 14, Cunha; 15, Domingos; 16, Emílio; 17, Espinola; 18, Ferreira; 19, Figueiredo; 20, Fraga; 21, Freitas; 22, Garcia; 23, Gomes; 24, Goulart; 25, Lima; 26, Machado; 27, Martins; 28, Medeiros; 29, Melo; 30, Mendes; 31, Mendonça; 32, Moura; 33, Nunes; 34, Oliveira; 35, Pacheco; 36, Pedras; 37, Pereira; 38, Picanço; 39, Pimentel; 40, Raposo; 41, Resendes; 42, Rocha; 43, Rodrigues; 44, Rosa; 45, Santos; 46, Silva; 47, Silveira; 48, Soares; 49, Sousa; 50, Vieira.

As previously stated, geographic distance is considered as one of the factors modulating the genetic structure of populations. The application of Malécot's model (Malécot, 1948) reveals that about 33% of the variation in kinship is explained by geographic distance (Fig. 5) and that this last variable has an important impact on the genetic structure of the archipelago, which is in accordance with the spatial autocorrelation analysis results obtained by Cabral *et al.* (2005) using data of surnames from the nine Azorean islands.

To visualize, on a geographic map, the trend data contained in the genetic distance matrix based on random isonymy, i.e. Monmonier's maximum-difference algorithm (Monmonier, 1973), were applied. As can be observed in Fig. 6, the most isolated island of the archipelago is Graciosa (a result similar to that obtained by Branco &
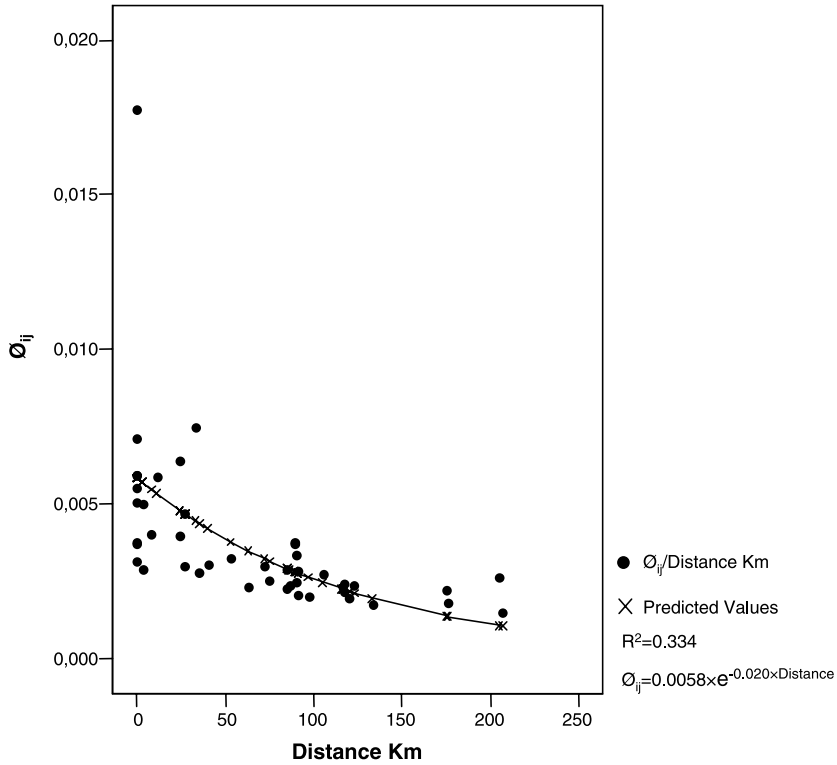
Fig. 5. Scatterplot of geographic distance (km) and *a priori* kinship. The line of predicted values using Malécot's isolation-by-distance model (Malécot, 1948) is shown.
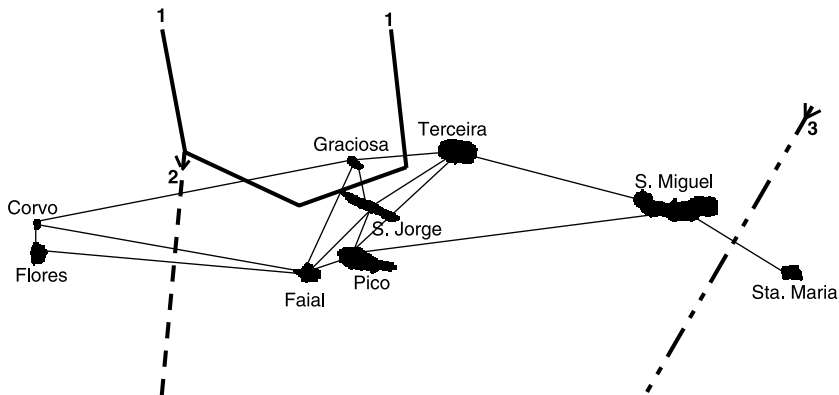

Fig. 6. Delaunay triangulation and the first three genetic barriers (1, 2 and 3) computed using a genetic distance based on a random isonymy matrix.

Mota-Vieira, 2005, using surnames from the nine Azorean islands), followed by the islands of the Western group and by Sta. Maria. This picture is highly supported by genetic data gathered so far: the analysis of the non-recombining region of the Y chromosome (NRY) haplogroups by groups of island (Montiel *et al.*, 2005) puts in evidence the differentiation of the Western group, which presents a significantly different distribution of haplogroups and a genetic diversity greater than that observed in the other groups of islands (Montiel *et al.*, 2005). Moreover, preliminary results obtained from the comparison by island of NRY markers also supports the differentiation of Sta. Maria with respect to S. Miguel (Bettencourt *et al.*, 2006).

When considering the islands grouped according to the three geographical clusters, AMOVA highlights the fact that there is no important variation amongst geographical groups. On the other hand, significant variation is observed amongst populations within groups, indicating that more variation exists between islands from the same geographical group than between geographical groups of islands, a result that is consistent with that obtained by graphical representations. Corroborating the results previously presented, if the islands are clustered according to the graphical analysis, a reduction of the within-population variation and an increase in among-groups variation is observed.

The results obtained in this study clearly demonstrate the need to perform a set of analyses when aiming to infer the sub-structure of a population. The interpretation of summary statistics, using limits deduced from different populations and markers, is clearly ambiguous and can result in missed interpretations. For example, for the Azores the exclusive analysis of values of F_{ST} has been used to suggest that the Azores archipelago have a homogeneous population (Branco & Mota-Vieira, 2005), a conclusion that is clearly not supported by a more exhaustive analysis. As is the case with other current statistics (such as means, for example), so summary statistics based on surnames should be evaluated using statistical tests and confidence intervals. In this work, it is found that the AMOVA formulation can constitute a good approach to attributing significance to F_{ST}. According to the results obtained, correspondence analysis is a good tool to identify which populations present marked differences in surname composition, with the advantage of having represented in the same plan both populations and surnames. However, it has a disadvantage for populations with a large number of surnames, such as the Azores, since it is impossible to consider all the surnames simultaneously in the analysis. Moreover, as with the representation of distance matrices using MDS, this technique does not account for the influence of geographic distance in genetic structure. If a significant influence of distance is detected, methods like Monmonier's maximum-difference algorithm, which provides a real graphic representation of genetic differences in geographic space, can be used successfully.

Taken globally, the analyses performed in this work, combined with the data of genetic markers gathered so far, indicate that the Azores archipelago is a sub-structured population, with sub-populations presenting low levels of diversity, such as Graciosa, whilst others present surprisingly high values of diversity, as is the case with the small island of Flores. In accordance with previous observations (Bettencourt *et al.*, 2006), these results further reinforce the need, whenever performing population genetic studies in this population, of a detailed genetic characterization of the

archipelago; using a representative sample of the nine islands of the archipelago is mandatory. Amongst other implications, the detailed genetic characterization will allow the identification of adequate sub-populations within the Azores to perform association mapping studies. Moreover, the data gathered will be important in the establishment of a specific forensic genetic database for the Azores archipelago.

Acknowledgment

References

Abade, A. (2003) *ISOnymy3 V. 3·0.18 – Software for Human Population Genetics Data (surname) Analysis*. Department of Anthropology, University of Coimbra, Portugal.

Bettencourt, C., Montiel, R., Santos, C., Prata, M. J., Aluja, M. P. & Lima, M. (2006) Diversity of maternal and paternal lineages in the geographic extremes of the Azores (Santa Maria and Flores Island): insights from mtDNA, Y-chromosome and surname data. In Amorim, A., Corte-Real, F. & Morling, N. (eds) *Progress in Forensic Genetics 11: Proceedings of the 21st International ISFG Congress* held in Ponta Delgada, The Azores, Portugal, 13th–16th September 2005. *International Congress Series* **1288**, 88–90.

Branco, C. & Mota-Vieira, L. (2005) Surnames in the Azores: Analysis of the isonymy structure. *Human Biology* **77**, 37–44.

Brassel, K. E. & Reif, D. (1979) A procedure to generate Thiessen polygons. *Geographical Analysis* **325**, 31–36.

Cabral, R., Branco, C. C., Costa, S., Caravello, G., Tasso, M., Peixoto, B. R. & Mota-Vieira, L. (2005) Geography of surnames in the Azores: specificity and spatial distribution analysis. *American Journal of Human Biology* **17**, 634–645.

Colantonio, S. E., Lasker, G. W., Kaplan, B. A. & Fuster, V. (2003) Use of surname models in human population biology: a review of recent developments. *Human Biology* **75**, 785–807.

Dahlberg, G. (1948) *Mathematical Methods for Population Genetics*. Interscience, Karger, Bâle, New York.

Excoffier, L., Smouse, P. & Quattro, J. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491.

Fernando, O., Mota, P., Lima, M., Silva, C., Montiel, R., Amorim, A. & Prata, M. J. (2005) The peopling of the Azores Islands (Portugal): data from the Y-chromosome. *Human Biology* **77**, 189–199.

Flint, J., Bond, J., Rees, D. C., Boyce, A. J., Roberts-Thomson, J. M., Excoffier, L. *et al.* (1999) Minisatellite mutational processes reduce F(st) estimates. *Human Genetics* **105**, 567–576.

Freedman, M. L., Reich, D., Penney, K. L., McDonald, G. J., Mignault, A. A., Patterson, N. *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nature Genetics* **36**, 388–393.

Gomes, F. (1997) *A ilha das Flores: da redescoberta à actualidade (subsídios para a sua história)*. Câmara Municipal de Lajes das Flores, Lajes das Flores, Azores.

Gregório, R. *Escravos e libertos da ilha Terceira na primeira metade do século XVI*. Livro de Homenagem ao Prof. Doutor Artur Teodoro de Matos. Centro de História de Além-Mar. Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (in press).

Harpending, H. C. & Ward, R. (1982) Chemical systematics and human populations. In Nitecki, M. (ed.) *Biochemical Aspects of Evolutionary Biology*. Chicago University Press, Chicago, pp. 213–256.

Hartl, D. & Clark, A. (1997) *Principles of Population Genetics*, 3rd edn. Sinauer Associates Inc. Publishers, Sunderland, MA.

Hey, J. & Machado, C. A. (2003) The study of structured populations – new hope for a difficult and divided science. *Nature Review Genetics* **4**, 535–543.

INE (Instituto Nacional de Estatística) (2001) *Recenseamento da População e da Habitação (R.A. dos Açores) – Censos 2001*. Instituto Nacional de Estatística, Portugal.

Jorde, L. (1980) The genetic structure of subdivided human populations. In Mielke, J. H. & Crawford, M. H. (eds) *Current Developments in Anthropological Genetics*, Vol. 1. Plenum Press, New York, pp. 135–208.

Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T. & Batzer, M. A. (2000) The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *American Journal of Human Genetics* **66**, 979–988.

McKeigue, P. M. (2005) Prospects for admixture mapping of complex traits. *American Journal of Human Genetics* **76**, 1–7.

Malécot, G. (1948) *Les Mathematiques de l'Heredite*. Masson, Paris.

Manni, F., Guérard, E. & Heyer, E. (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by using Monmonier's algorithm. *Human Biology* **76**, 173–190.

Marchini, J., Cardon, L. R., Phillips, M. S. & Donnelly, P. (2004) The effects of human population structure on large genetic association studies. *Nature Genetics* **36**, 512–517.

Mendonça, L. (1996) *História dos Açores – Visão geral (sécs. XV–XIX)*. Centro de Apoio Tecnológico à Educação, PontaDelgada, Azores.

Merelim, P. (1966) *Os Hebraicos na Ilha Terceira. Angra do Heroísmo, Açores*. Separata da Atlântida.

Mesquita, H. Escravos em Angra no século XVII (uma abordagem a partir dos registos paroquiais). *Arquipélago, História, II série*. vol. IX. (in press).

Monmonier, M. (1973) Maximum-difference barriers: an alternative numerical regionalization method. *Geographical Analysis* **3**, 245–261.

Montiel, R., Bettencourt, C., Silva, C., Santos, C., Prata, M. J. & Lima, M. (2005) Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores archipelago (Portugal). *Annals of Human Genetics* **69**, 1–10.

Nei, M. (1987) *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Neto, D., Montiel, M., Bettencourt, C., Santos, C., Prata, M. J. & Lima, M. The African contribution to the present-day population of the Azores Islands (Portugal): Analysis of the Y chromosome haplogroup E. *American Journal of Human Biology* (in press).

Relethford, J. (1988) Estimation of kinship and genetic distance from surnames. *Human Biology* **60**, 475–492.

Santos, C., Abade, A., Cantons, J., Mayer, F., Aluja, M. P. & Lima, M. (2005) Genetic structure of Flores island (Azores, Portugal) in the 19th century and in the present day: evidence from surname analysis. *Human Biology* **77**, 317–341.

Santos, C., Lima, M., Montiel, R., Angles, N., Pires, L., Abade, A. & Aluja, M. P. (2003) Genetic structure and origins of peopling in the Azores Islands (Portugal): the view from mtDNA. *Annals of Human Genetics* **67**, 433–456.

Schneider, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN ver. 2·000: A Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Spinola, H., Brehm, A., Bettencourt, B., Middleton, D. & Bruges-Armas, J. (2005) HLA class I and II polymorphisms in Azores show different settlements in Oriental and Central islands. *Tissue Antigens* **66**, 217–230.

SPSS (2003) *SPSS for Windows – Release 12·0.1*. SPSS Inc., Chicago.

Tomaz, J. (1966–1969) Lista dos escravos da ilha das Flores, tirade dos livros de registo de nascimento, casamento e óbito existents nas conservatories do registo civil de Lajes e Santa Cruz. *Boletim do Núcleo Cultural da Horta* **4**, 105–114.

Weir, B. S. (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data*. Sinauer Assoc. Inc., Sunderland, MA, USA.

Weir, B. S. & Cockerham, C. C. (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics* **15**, 323–354.

Wright, S. (1984) *Evolution and the Genetics of Populations: Variability Within and Among Natural Populations*. University Press, Chicago.

Ziv, E. & Burchard, E. G. (2003) Human population structure and genetic association studies. *Pharmacogenomics* **4**, 431–441.