Genetic structure and population history of Ireland: a comparison of blood group and anthropometric analyses

J. H. RELETHFORD
Department of Anthropology, State University of New York College at Oneonta, NY

Received 19 January 1983

Summary. Population structure and history may be studied on a local or a regional level. This paper examines the regional population structure of the Republic of Ireland and Northern Ireland with respect to population history and demographic processes. Blood-group and anthropometric data obtained from the literature are analysed. The blood-group data consist of ABO and Rhesus gene frequencies for 32 counties and the Aran Islands. Anthropometric data consist of summary statistics for 15 variables collected from 19 regions. The degree and pattern of population differentiation is assessed using new methods of population-structure analysis. Both blood group and anthropometric analyses show a west-east division of populations corresponding to the known history of inhabitation of Ireland, where successive waves of immigrants pushed earlier populations further west. In both analyses there were two deviations to this basic pattern: the Aran Islands and the midlands. In both cases, alternative historical explanations are examined. The genetic relationship of the Aran Islands to the rest of Ireland and England appears to be due to English admixture following the garrisoning of soldiers several centuries ago. The genetic position of the midlands is more complex, but suggests the effects of early Viking inhabitation. These findings are related to studies of the local, rather than regional, population structure of Ireland.

1. Introduction

In recent years, studies of human population structure have allowed the investigation of cultural and biological interrelationships in terms of the analysis of genetic effects of mating structure. Mating probabilities can be influenced by the composition of a population, preferences and/or avoidances of consanguineous marriage, and barriers to gene-flow, such as geographic distance and social class. Historically oriented studies have allowed changing patterns of population structure to be assessed. Here, known migration patterns, population movements, changes in population size, and general aspects of demographic change are shown to affect the genetic structure of regional populations. Studies of regional, rather than local, population structure have been done for a variety of European countries, including Belgium (Dodinval 1970), Switzerland (Morton, Klein, Hussels, Dodinval, Todorov, Lew and Yee 1973), Finland (Workman, Mielke and Nevanlinna 1976), Italy (Zanardi, Dell'Acqua, Menini and Barrai 1977), Iceland (Jorde, Ericksson, Morgan and Workman 1982), and England (Roberts, Jorde and Mitchell 1982).

Ireland (including the Republic of Ireland and Northern Ireland) is a country whose unique history and demography have long interested social scientists (Kennedy 1973). To date, research on genetic structure and population history of Ireland have followed two different paths. First, there have been a number of large regional studies focusing on basic patterns of variation with respect to history on a trait-by-trait basis. These studies have dealt with both anthropometrics (Hooton and Dupertuis 1955) and blood groups (Hackett, Dawson and Dawson 1956, Hackett and Dawson 1958, Dawson 1964, Tills 1977 a, Tills, Teesdale and Mourant 1977). These studies have been generally descriptive, univariate, and not integrated in the context of population structure theory. The second type of study has dealt with local population structure,
focusing on rural populations in western Ireland (Relethford 1980 b, Relethford, Lees and Crawford 1980, 1981, Lees and Relethford 1982) and west-coast isolates (Relethford 1982). These studies have dealt with the methods and theories of population structure, but describe only those patterns of interest on a local level. It is expected that the cultural and demographic processes operating on genetic structure will be different on a regional level.

The purpose of this paper is to investigate the regional population structure of Ireland using newly developed methods of population structure analysis. The main emphasis of this analysis is to relate genetic structure to population history. Results are compared with other European populations and with studies of local population structure in Ireland.

2. Materials

Blood-group and anthropometric data are available in the published literature for a regional analysis of Ireland (The term ‘Ireland’ as used in this paper refers to the Republic of Ireland and Northern Ireland collectively). Blood-group data are limited to the ABO and Rhesus loci. While several studies have incorporated additional loci, they are limited to only a few populations or the entire island as a whole (e.g., Sunderland, Tills, Bouloux and Doyle 1973, Walter and Palsson 1973, Tills 1977 a, Tills et al. 1977). In order to include a large number of populations representing the entire region, all blood-group analyses had to be confined to the two loci. While a large number of additional loci would be preferred, the use of only the two loci is still highly valuable, as shown by a number of studies on population structure where only these loci (or even just one) were available (e.g. Imaizumi and Morton 1969, Dodinval 1970, Morton et al. 1973, Zanardi et al. 1977, Jorde et al. 1982). In any case, the usefulness of any data set rests on the ability to interpret the results in a meaningful manner.

The basic unit of analysis for blood-group study in this paper is the county. The Republic of Ireland has 26 counties and Northern Ireland has 6. In addition to these 32 counties, the Aran Islands have been treated as a separate population. The Aran Islands consist of three small islands in Galway Bay off the west coast of County Clare and County Galway. The islands have been treated as a separate population due to their geographic isolation and previous reports of suggested genetic divergence from the mainland of Ireland (Hooton and Dupertuis 1955, Hackett and Folan 1958). The locations of all the counties are shown in figure 1. Three studies were used to provide the gene frequencies needed for analysis. Data for the 26 counties in the Republic of Ireland were taken from Dawson (1964), which lists data for 117287 adult men and women. This sample is approximately 5.5% of the total adult population of the Republic of Ireland at the time of sampling. Gene frequencies for the Aran islands were taken from Hackett and Folan’s (1958) study of 229 people. Gene frequencies for the six counties in Northern Ireland were taken from Tills et al. (1977), who compiled the earlier data of Kopeć (1970). This sample consists of 27861 men and women for the ABO locus and 32799 men and women for the Rhesus locus. Overall, the sample sizes from these studies are high, compensating to some extent for the limited number of loci for analysis.

Anthropometric data are also available on a regional basis. These data were taken from Hooton and Dupertiuis (1955), and were collected originally as part of
Figure 1. Location of Populations used in the blood-group analysis.

1, Aran Islands; 2, Mayo; 3, Galway; 4, Clare; 5, Limerick; 6, Kerry; 7, Sligo; 8, Roscommon; 9, Tipperary; 10, Cork; 11, Donegal; 12, Leitrim; 13, Longford; 14, Westmeath; 15, Offaly; 16, Leix; 17 Kilkenny; 18, Waterford; 19, Cavan; 20, Meath; 21, Kildare; 22, Carlow; 23, Monaghan; 24, Louth; 25, Dublin; 26, Wicklow; 27, Wexford; 28, Fermanagh; 29, Tyrone; 30, Derry; 31, Armagh; 32, Antrim; 33, Down. The Aran Islands (1) are part of County Galway (3) but are treated separately here. Counties numbered 28–33 are in Northern Ireland.

the Harvard Anthropological Survey of Ireland in the 1930s. The data are reported in ‘county subgroups’, which were defined by Hooton and Dupertuis as single counties, multiple counties, or subgroups composed of parts of various counties. Nineteen county subgroups were used by Hooton and Dupertuis, and are shown in figure 2. Since the units of analysis are not the same for blood-groups and anthropometric data, certain analytical techniques, such as correlation analysis and rotational fitting, could not be used, and comparison is limited to visual inspection of patterns of variation. While a portion of the original data have been analysed in previous papers (Relethford 1980 a, b, Relethford, Lees and Crawford 1980, Relethford and Lees 1981, Relethford et al. 1981, Lees and Relethford 1982, Relethford and Lees 1983), only summary data are available for the entire population of Ireland. Also, these summary data are limited to adult males, since data on adult females were collected only in a few western counties. The summary data consists of means and standard deviations for 15 anthropometric variables: stature, biacromial breadth, chest breadth, chest depth, sitting height, head length, head breadth, head circumference, head height, bizygomatic breadth, bigonial breadth, minimum frontal diameter, upper facial height, nose length, and nose breadth. Since environmental variation may contribute to among group variation in body size, all analyses were repeated using only the ten craniofacial variables, since they are often quite sensitive in detecting genetic affinities (Friedlaender 1975, Relethford et al. 1980).
3. Methods

Population affinities were assessed using measures of genetic structure and microdifferentiation. For blood-group gene-frequency analyses, the R matrix technique of Harpending and Jenkins (1973) was used. Here, a matrix of standardized genetic variances and covariances among populations is computed for the kth allele as

$$r_{ik} = \frac{(P_{ik} - \bar{P}_k) (P_{jk} - \bar{P}_k)}{\bar{P}_k (1 - \bar{P}_k)}$$

where $P_{ik}$ and $P_{jk}$ are the gene frequencies for allele $k$ in populations $i$ and $j$, and $\bar{P}_k$ is the weighted mean gene frequency for allele $k$ over all populations. Each element in the R matrix is then computed as the mean value over all alleles. Each element in the matrix represents divergence or co-divergence from the contemporary regional gene-pool, and has also been referred to as 'conditional kinship' (Morton, Yee, Harris and Lew 1971). Positive $r_{ij}$ values indicate a pair of populations more similar...
than expected at random, while a negative $r_{ij}$ value indicates the opposite. A mean value of conditional kinship within groups, $R_{ST}$, may be computed as the weighted mean of all $r_{ii}$ values. Additionally, this value may be computed after adjustment for sampling bias (Workman, Harpending, Lalouel, Lynch, Niswander and Singleton, 1973). The $R_{ST}$ value is an estimate of the average divergence from a contemporary gene-pool. As such, $R_{ST}$ is similar to Wright's $F_{ST}$, except that the latter refers to divergence from a founding, and not contemporary, gene-pool.

Since visual inspection of a 33 by 33 matrix of conditional kinship values precludes any comprehensive synthesis, this information must be represented in a more usable manner. Here, the method of Harpending and Jenkins (1973) is used, where the eigenvalues and eigenvectors of the R matrix are extracted using principal components analysis. The eigenvectors represent the positions of the populations along scaled axes which have been extracted in descending order of explained variation. As such, often only a few axes are needed to provide an accurate representation of genetic variation in a multidimensional space. Patterns of variation among alleles can also be investigated in order to detect which alleles contribute the most (or least) to specific patterns of among-group variation (Harpending and Jenkins 1973).

Anthropometric variation is also assessed using a measure of population differentiation. Here, the Euclidean distance between each pair of populations is computed as

$$d_{ij}^2 = \frac{\sum (\bar{X}_{ik} - \bar{X}_{jk})^2}{k}$$

where $\bar{X}_{ik}$ and $\bar{X}_{jk}$ are the standardized means for variable k for populations i and j. Euclidean distance is directly related to conditional kinship as shown by Morton et al. (1971) and Relethford (1980 a). Thus, Euclidean distance can be used for genetic inference assuming that environmental and sampling effects are either minor or distributed at random across all populations. A measure such as Mahalanobis' $D^2$ would be preferred since it controls for intercorrelation and redundancy of information, but is not possible to use given summary data. An eigenvectorial plot can be produced from the distance matrix using Gower's (1966) method of converting distances into similarity measures and then using the standard principal components method.

4. Results

The average amount of divergence from the regional gene-pool ($R_{ST}$) is 0.00113 for the ABO system and 0.00057 for the Rh system, giving an overall value averaged over five alleles of $R_{ST} = 0.00091$. After correction for sampling bias, the estimates are similar: ABO = 0.00102, Rh = 0.00047, total = 0.00080. The overall level of genetic microdifferentiation is greater than values for regional studies in Scotland (0.00030), England (0.00034) and Czechoslovakia (0.00031), similar to values from France (0.00087) and Sweden (0.00078), and less than values from Finland (0.00220) and Ireland (0.00123) (comparative values were taken from Jorde 1980). Overall, the $R_{ST}$ value for Ireland is in the range expected. The values reported in the present study are lower than those given by Tills (1977 b) in his analysis of gene-frequency variation among Irish counties (ABO = 0.01200, Rh = 0.01875, combined = 0.01538). In that
J. H. Relethford

study, however, the sample sizes were extremely small, with a total sample of 1793 for the ABO locus and 1896 for the Rh locus (Tills et al. 1977).

While the average value ($R_{ST}$) is low, some of the individual counties show great divergence from the regional mean. The three largest values of $r_{ii}$ are for County Clare (0.00814), the Aran Islands (0.00691) and County Cork (0.00403). By contrast, the three lowest values are for County Meath (0.00007), County Derry (0.00011) and County Kildare (0.00012).

Individual variation in genetic divergence can be shown more clearly by an examination of the eigenstructure of the $R$ matrix. Figure 3 is a plot of the first two eigenvectors extracted from the $R$ matrix, with eigenvector 1 accounting for 69.5% of the total variation, and eigenvector 2 accounting for 17.2%. Each axis in figure 3 has been scaled by the square root of its corresponding eigenvalue in order to provide a more accurate picture of relative dispersion (Harpending and Jenkins 1973).

![Figure 3. Plot of the first two eigenvectors of the R matrix based on blood groups. Each axis is scaled by the square root of the corresponding eigenvalue.](image)

The first axis separates populations along a west–east division for the most part. The populations grouping to the extreme left are those located on or adjacent to the west coast of Ireland: Clare (4), Cork (10), Roscommon (8), Galway (3), Kerry (6) and Mayo (2). Those populations plotting to the immediate left of centre are those in the midlands. For the most part, populations plotting to the right of centre are those which are on, or adjacent to, the east coast. Several of these populations are markedly separated from the rest, including Armagh (31), Wexford (27), Wicklow (26) and Fermanagh (28). Finally, the Aran Islands (1) plot to the extreme right, in total opposition to the general west–east pattern. Taken as a whole, the Northern Irish populations (28–33) tend to plot further to the right than their neighbours in the Republic.

Overall, the west–east pattern shown in figure 3 could have two interpretations. First, that geographic distance acts to subdivide the entire population according to the isolation by distance model, such that the further two populations are in geographic space, the more genetically dissimilar they will be (Jorde 1980, among
others). The second interpretation is that repeated waves of immigrants in the past have pushed previous populations further westward (Dawson 1964). The separation of populations along the second axis weakens the isolation-by-distance argument, at least as applied to isotropic migration. Instead of a north-south separation, the second axis acts to separate the populations of Roscommon (8) and Longford (13). The third axis (not shown) accounts for the remaining 13·2% of the total variation, and acts only to separate County Clare (4) even more. Plots of variation among alleles (not shown) show that the basic west-east pattern represents a contrast of high frequencies of the 0 allele in the west versus high frequencies of the A allele in the east. This pattern is augmented to some extent by high Rh(D) frequencies in the west and high Rh(d) frequencies in the east. The distinctiveness of Longford and Roscommon is associated with high frequencies of B and Rh(D).

Thus, the general picture of the genetic map is a west–east division with two noticeable deviations from this pattern: (1) the divergence of Longford and Roscommon from other mid-land populations, and (2) the extreme divergence of the Aran Islands from the west–coast populations. Alternative explanations are explored in the discussion below.

The first two eigenvectors of the $d^2$ matrix based on anthropometric data are shown in figure 4, where both axes have been scaled by the square root of their corresponding eigenvalues. Axis 1 accounts for 44·6% of the total among-group variation and, excluding the Aran Islands (1) and the Longford–Westmeath (11)
subgroups, shows a distinct west-east separation. The west-coast populations (2, 3, 4, 5 and 10) are clustered on the extreme left, while populations on or adjacent to the east coast (12, 15, 16, 18 and 19) are clustered on the right. Midland populations tend to plot near the centre of the picture. The second axis accounts for 32.6% of the total variation, and acts to separate the Aran Islands and the Longford–Westmeath subgroups. The third axis (not shown) accounts for only 9.8% of the total variation and acts to separate these two subgroups even more. An analysis using only craniofacial variables gave essentially the same results, suggesting that variation in body measurements does not enhance or obscure the basic relationships. Overall, the results from the anthropometric distance analysis supports the idea of a basic west–east separation, along with the distinctiveness of the Aran Islands and certain mid-land populations.

5. Discussion

While there are expected differences between the blood-group and anthropometric analyses, there are several common features which may be examined in light of the population history of Ireland. As indicated above, the separation of populations along a west–east division suggests a correspondence of genetic affinities with geographic distance and/or historic patterns of population movement. Earlier work (Hooton and Dupertuis 1955, Dawson 1964) has shown this west–east pattern on a univariate basis, but did not assess alternative hypotheses. Nor did they combine results in a multivariate way. While geographic distance could explain the basic west–east gradient, there is no north–south gradient evident in the present results. Historic population movements can be used to explain the overall trends noted. The invasion model of early Irish occupation (until about 2000 B.P.) states that the original Celtic inhabitants came in four succesive ‘waves’, whereby each new wave pushed earlier inhabitants further westward (O’Rahilly 1946). Similar arguments have been used to explain the west–east gradient on a univariate basis (Hooton and Dupertuis 1955, Dawson 1964).

Western and eastern regional differences were probably enhanced by additional historic population movements. Some events were probably minimal in effect, such as the Anglo-Norman invasion of 1169 which involved few men (Sunderland et al. 1973). After this time, however, there was an increase in Anglo-Norman invasion and settlement until about 1300, occurring for the most part in eastern Ireland. Subsequent invasions were small in number until the influx of Scottish and English Protestants into Northern Ireland and the eastern portions of the Republic as the result of the Articles of Plantation of James I in 1609 (Hackett, Dawson and Dawson 1956).

The patterns of biological variation shown here do not rule out geographic distance as a factor influencing regional population structure, merely isotropic migration with respect to geographic distance. There are several studies which demonstrate that west–east migration was common during the last century, due mostly to increasing rural–urban migration with Dublin as a prime target of many migrants (Geary and Hughes 1970, Gillman 1970, Kennedy 1973). Such migration seems unidirectional in nature. As such, recent migration might have acted to counter, and not enhance, previous patterns of population movement associated with inhabitation of western regions.

This recent west–east migration might explain the fact that the anthropometric map shows greater distinctiveness of the west–coast populations than the blood-
Genetic structure of Ireland

Previous studies have suggested that anthropometrics are less susceptible to recent events, and tend to show long-term patterns to a greater extent than serological traits (see Relethford and Lees 1982 for a review of these studies). Another possibility is that too few loci are available for the blood-group analyses to show as clear a pattern.

The distinctiveness of the Aran Islands, shown by both blood-group and anthropometric analyses, can be explained in terms of historical events. Original inhabitation of the islands seems to have resulted from continued westward movements of the Irish population (Messenger 1969). In the late sixteenth century the island took on strategic value and English soldiers were frequently garrisoned. The first occurrence was in 1587 when 20 soldiers were placed on the islands at the command of Queen Elizabeth. Following this, there were several times when English soldiers were sent to the islands, often in groups of several hundred, until the late seventeenth century. Many soldiers stayed on the islands to marry local women. Thus, a repeated influx of English men would have had considerable genetic impact, particularly since such events occurred when the local population size was low (Haddon and Browne 1893, Hackett and Folan 1958).

English affinities of the Aran Islanders has been suggested in several earlier works (Hooton and Dupertuis 1955, Hackett and Folan 1958) and is confirmed by the present study. Specific affinities are difficult to determine from anthropometric data. The blood-group data show the Aran Islands are genetically most similar to the east-coast populations of Ireland, as expected from a model of English admixture. Further confirmation is provided by an examination of the R matrix based on four populations: the Aran Islands, the Republic of Ireland, Northern Ireland and England. English gene frequencies were taken from Race and Sanger (1968). The first two axes of the resultant eigenvectorial plot is shown in figure 5. The first axis accounts for 93.9% of the total variation, and clearly illustrates the relationship of the Aran Islands to England. Taking the Aran Islands as a dihybrid population, a rough estimate of the relative parental contributions can be made. First, the R matrix is converted into a distance matrix following the transformation suggested by Harpending and Jenkins (1973) \( d_{ij}^2 = r_{ii} + r_{jj} - 2r_{ij} \). Second, the distances between the hybrid and parental populations can be used to estimate admixture. Using the method of Morton, Dick, Allan, Izatt, Hill and Yee (1977), the estimates of admixture are 41% Irish and 59% English. The method of Lees and

![Figure 5. Plot of the first two eigenvectors of the R matrix comparing the Aran Islands to other populations. Each axis is scaled by the square root of the corresponding eigenvalue.](image)
Relethford (1978) yield similar estimates of 43% Irish and 57% English. Again, these estimates are crude first-order approximations, but they do suggest an interesting population for further admixture studies.

Both blood-group and anthropometric analyses suggest that genetic distinctiveness of several midland populations. For the blood-group analysis, the counties of Longford (13) and Roscommon (8) are separate from nearby populations. For the anthropometric analysis, the county subgroup of Longford–Westmeath (11) is quite distinct. Hooton and Dupertuis (1955) noted the distinct nature of this subgroup and suggested two explanations: (1) age-related variation in that this subgroup was on average the youngest, and (2) effects of early Viking inhabitation in the area. In their original analysis, Hooton and Dupertuis noted that the young average age of men in Longford–Westmeath could account for some of the group's distinctiveness. In the present study, however, this separation remains even after variables most affected by age are removed. The hypothesis of Viking contact has historical validity. Early Norse invasions of Ireland began in the late ninth century, and the major settlements sprang up in the coastal ports of Wexford, Waterford, Dublin, Cork and Limerick. The only significant inland occupation of the Vikings seems to have occurred around Lough Ree, a lake bordering Longford, Roscommon and Westmeath. This settlement seems to have resulted from expansion of the Limerick Norse up the Shannon River (Hooton and Dupertuis 1955, Jones 1968, Orme 1970). In fact, the Gaelic term 'longphort' (the Gaelic equivalent of Longford) means a Norse fortification (Orme 1970). The relationship of history to naming of the county is not known, however. That other Norse settlements do not show any distinction in blood-group or anthropometric analyses may be due to the fact that these were coastal ports, and as such experienced greater immigration in later times. That is, Norse colonization is a slight factor relative to later historical events. The Norse may have had a greater genetic effect in the midlands.

Even though the blood-group and anthropometric analyses are similar with respect to Longford, there are still several differences which must be considered when evaluating the Viking hypothesis. Roscommon is divergent along with Longford in the blood-group analysis, but the county subgroup containing Roscommon (8) in the anthropometric analysis is not. Likewise, Westmeath (14) is not divergent in the blood-group analysis. The probable cause of these differences lies in the fact that the two analyses use different sampling units. For example, Roscommon is combined with the eastern part of Galway in the anthropometric analysis. When different regions are pooled, the overall relationship matrix may be changed. In any case, the present paper can offer no conclusions, but only suggestions for further studies in the Irish midlands, particularly with a much larger set of variables.

In sum, the present paper shows that the regional population structure of Ireland has been affected by a series of historical population movements. The basic pattern of a west–east gradient suggests a series of genetically different immigrants over a long period of time. Deviations from this pattern are noticed in the Aran Islands and the mid-lands, both of which have historical explanations. A final question to be examined is: how does regional population structure compare with local population structure?

Two previous studies of local population structure allow some comparison of levels. Both of these studies are based on local variation in areas of western Ireland: (1) rural mainland western Ireland during the 1930s (Relethford 1980b, Relethford,
Lees and Crawford 1980, 1981), and (2) west-coast islands and isolates in the 1890s (Relethford 1982). In both studies, geographic distance was found to be a major determinant of population structure, with good fits to the expected isolation by distance model. However, the influence of geographic distance was much stronger in the west coast isolates. On the mainland, other factors, such as population size, influenced the degree of migration and genetic similarity. Thus, in the most isolated areas geographic distance remains the single most powerful factor influencing population structure. In a mainland situation geographic distance remains important, but is also supplemented by factors such as the 'pull' of larger populations in accordance with gravity models of migration. Finally, on a regional level, previous historic patterns hold, although continued patterns of unidirectional migration toward Dublin may soon act to counter the basic pattern of variation.

It is clear that much work remains to be done in Irish population structure, both on a local and regional level. The results presented here may be taken less as definite conclusions, and more as suggestions for future research. The uniqueness of Irish history and demography offers a variety of prospects in human population biology.

References


O'Rahilly, T. F., 1946, Early Irish History and Mythology (Dublin: Institute for Advanced Studies).


Tills, D., 1977 b, The use of the FST statistic of Wright for estimating the effects of genetic drift, selection and migration in populations, with special reference to Ireland. Human Heredity, 27, 153–159


Address correspondence to: Dr John H. Relethford, Department of Anthropology, State University of New York at Oneonta, Oneonta, NY 13820, USA.
