SIR JOHN RHYS MEMORIAL LECTURE

The Genetics of Celtic Populations

WALTER F. BODMER
Imperial Cancer Research Fund, London

Introduction

I am honoured to have been invited to give the Rhys Lecture to the British Academy, and that you have entrusted a scientist with this task. This is no doubt with the aim of helping to create a bridge between the humanities and the sciences, at least in this area where it is especially appropriate. But I am also embarrassed to give this Lecture following so many real experts on the Celtic people, their culture and their history. My main contact with Wales has been that of many holidays with happy memories, a brief period of evacuation to Prestatyn at the beginning of the second world war when I was too young to go to school to learn Welsh, but when my older brother was introduced to the language having only just learnt English, and having a wife whose middle name is Gwynaeth, a name her mother chose from a Criccieth tombstone. Perhaps, however, I may convince you that I might share some middle European genes with those brought by the ancestors of the Celts to Wales.

My aim is to explain how the study of genetic variation can be used to define populations on a statistical basis and to elucidate their movements and intermingling, and to emphasize the way that such information might be integrated with the historical record and with archaeological and anthropological investigations. My title is clearly far too ambitious, and to start it with the definite article must be considered little less than presumptuous. A more appropriate title might be 'an attempt to suggest approaches to defining the origin of Welsh and perhaps other Celtic populations on the basis of genetic studies'. Brevity is my only excuse for presumption.

Figure 1. The identical cricketing Bedser twins, Alec and Eric as babies, a few years old, and in their seventies. (Courtesy of Alec Bedser.)
Individual Differences and Racial Classifications

Apart from identical twins, everybody looks different from everyone else. The fact that identical twins really are almost indistinguishable at all ages, shows that these differences by which we recognize each other, are largely genetically determined. That is because identical twins are essentially the only pairs of individuals that are identical genetically (Figure 1). The range of individual differences can be enormous. Consider, for example, the contrast between the tall, white, distinguished British Anthropologist, Colin Turnbull and one of the Pygmies whom he studied with such insight. Physical features such as colour of the skin, of the hair, and of the eyes, height, body build and facial features have traditionally been used to define so called racial characteristics. It is, for example, often said that red hair is a particular feature of Celtic populations, including the Welsh, and Deo Cassius apparently described the Iceni Queen, Boudicca as '... huge of frame and terrifying of aspect and with a harsh voice. A great mass of bright red hair fell to her knees'. Red hair is undoubtedly genetically determined but is not, of course, a universal characteristic of Celtic populations. It may occur with higher frequency in some such populations, especially in Ireland, Scotland and Wales, and perhaps parts of Scandinavia, and so it is the frequency of this trait which can help to define the populations in which it is more commonly found. Populations, or races, can be defined by a combination of genetic and cultural traits including, of course, especially language. My emphasis here is on the objective definition of populations by the distribution of the frequencies of well defined genetic traits. For this, we must look to objectively defined chemical differences whose inheritance follows the simple patterns originally defined by the Moravian Monk, Gregor Mendel, and which follow the behaviour of the chromosomes that are the carriers of the genes.

Some Genetic Background

When sperm meets egg the equal inherited, or genetic, contributions of a male and a female unite to form the fertilised egg, a single cell from which all the hundred million million cells of the human body are derived by simple cell division. When a cell divides its chromosomes also divide and are passed on faithfully to the daughter cells. At the time of division the chromosomes, literally coloured bodies, can be seen down the microscope (Figure 2). There are forty-six human chromosomes, twenty-two identical looking pairs and the X and Y which determine sex, XX for a female and XY for a male. One member of each pair comes from the father, through
the sperm, and the other from the mother, through the egg. It is the behaviour of the chromosomes when sperm and egg are formed which determine the simple patterns of inheritance that ensure, for example, that males and females are produced in equal proportion. The chromosomes carry the genetic information, the genes.

**ABO Blood Types**

If you have ever given blood as a blood donor, and in particular if you have ever had a blood transfusion, you will know that it is important for the blood of donor and recipient to be typed and matched. The differences that are being matched are inherited chemical differences on the surface of red blood cells and were first identified by Karl Landsteiner in 1900. These are the ABO types, which are still the main ones that need to be matched for a blood transfusion. They are identified by a clumping test (Figure 3), such that if red cells are clumped with reagent anti-A, then they type A, while if clumped by reagent anti-B, they type B, creating four...
Figure 3. The ABO blood groups. The reaction patterns of anti-A and anti-B define the four types AB, B, A and O. (After Bodmer and Cavalli-Sforza 1976 page 215.)

types, AB, B, A, and O. These are simply inherited, such that if an AB individual marries someone who is O, then on average half their children will be A and the other half B. Thus, A, B and O are different versions of a gene, carried on chromosome 9. The AB individual has the A version on the chromosome 9 from one parent and the B on that from the other, and passes each on with equal probability to sperm, or egg. That is the essence of the basic laws of inheritance by which traditionally a particular trait has been defined as being clearly genetically determined.

The distribution of ABO types in different populations was the first example of the use of a genetic marker, a clearly inherited difference, for the definition and identification of populations. Because of the simplicity of the test and its importance medically for blood transfusion, many tens of thousands of people have been typed in different populations throughout the world.

The frequency of type B, for example, varies from less than 10 per cent in Northern European populations, such as in England, to nearly 50 per
Figure 4. World map of the distribution of the blood group gene B in aboriginal populations. The map is computer generated using data from many different populations. The heaviest shading indicates the highest frequency (gene frequency 0.225 to 0.275) with declining frequencies to the lowest shading (gene frequencies from 0 to 0.025). (From Cavalli-Sforza and Bodmer 1976 page 566.)

cent in Asian populations, such as in India. A world map of the distribution of the blood group gene B in aboriginal populations, computer generated using data from many different populations (Figure 4) shows clear gradients, of decreasing frequency moving westwards from Asia through Europe and to the north. This is, for example, consistent with the high frequency of type B found in a number of Central European populations of gypsies, and the fact that the gypsies are known to have originated from India. The much lower B frequency of Welsh gypsies, almost indistinguishable from that of the native population, suggests that by the time the gypsies had migrated there, there had been considerable admixture with local populations.

Because of the need to set up blood transfusion in a systematic way, the second world war provided an unprecedented opportunity for collecting large scale data on the frequencies of the ABO bloodgroups throughout Britain. One of the early reports (Fisher and Taylor 1940), pointed out that the frequency of type A increased markedly as one went south from Scotland, with a frequency of just over 30 per cent, through to Southern England and especially East Anglia, with a frequency of more
Figure 5. The frequency of blood group A gene in Wales. (From Morgan Watkin in Harper and Sunderland 1986 page 122.)

than 45 per cent. The relatively high frequency in East Anglia, for example, is consistent with Scandinavian colonization and their relatively high A frequency, while the lower frequency in the north of Scotland, which matches that in Iceland, Fisher thought may reflect a slightly different origin for the Viking invasions in the north and the west. However, this may, I believe, be more likely to reflect the pre-neolithic origins of the most northern populations in Scotland.

Extensive data on the ABO distribution have been collected in Wales (Figure 5). (Morgan Watkin in Harper and Sunderland 1986.) To limit the typing as far as possible to the genuinely Welsh population, data were obtained only from those individuals who had Welsh surnames (in the case of married women, Welsh maiden names), and whose place of birth was in Wales. The most striking feature of the distribution is the identification of ‘Little England beyond Wales’ in Southern Pembrokeshire by its relatively high A frequency. This agrees with the historical record showing that in early Norman times, around AD 1100, Flemish colonies were established there. The relatively low A frequency to the north may reflect
Figure 6. The distribution of the frequency of the HLA-A1 gene in various populations throughout the world. Each circle represents a population sampled from that position on the map and the size of the darkened sector is proportional to the gene frequency. (Courtesy of T. Gujabori from Tsuji et al. 1992.)

a lesser English admixture there, (as in the North of Scotland) while the relatively high frequency to the east presumably reflects the boundary with England. Clearly, however, more data especially using more genetic markers showing consistent patterns of variation, are needed to support these suggestions.

Tissue Types: The HLA System

While it is possible to match people by their ABO types, and to some extent other blood group differences, for blood transfusion, matching for tissue transplantation such as kidney or bone marrow grafts, is a very much more difficult problem. This is because the genetic differences that determine graft rejection are much more extensive and complex. There are more than two hundred individual types that fall into six categories, or genes A, B, C, DR, DQ and DP. Each type corresponds to a different version of one of those six genes and the total number of combinations possible is many millions. This system, called HLA, is thus extraordinarily informative for the definition of populations, and it is particularly because the six genes are close together on a chromosome, number 6, so that they are mostly inherited en bloc that it is at least possible to match within families. The pattern of inheritance ensures that on average there is a chance of a quarter that an individual’s brother or sister will have the
identical HLA type. As part of a recent international collaboration, the eleventh in a series, more than eighteen thousand individuals from more than seventy different populations throughout the world were HLA typed.

The type A1 has been known for some time to be characteristic of Caucasoid populations. Its worldwide distribution (Figure 6) shows very clearly this feature of its distribution, emphasizing that it is as frequent in European, namely western Caucasoids, as in Indian, namely eastern Caucasoid populations. Caucasoid admixture is, of course, clearly shown in the populations of the East Coast of America. More interesting, however, is the indication of a gradient in the frequency of A1 in China, with high frequencies in the North, in Mongolia, and much lower frequencies in the south. This pattern of distribution is paralleled by other Caucasoid related genes and suggests a distribution that parallels the Mongolian invasions from north to south in China, suggesting that the Mongolians in fact brought Caucasoid genes down into China. It is surely no longer appropriate therefore to refer to Oriental populations as Mongoloids!

The world distribution of A31 (Figure 7) is quite different and suggests a centre of origin for this type somewhere around Mongolia. Its relatively high frequency in American Indians may be explained by migration from there across the Behring Straits. Perhaps my A31 reflects an occasional migrant westwards!

The patterns of frequencies of different HLA types in different
populations can be used to characterize these populations, as indeed can those of any other set of genetic markers. The HLA System, however, because of its extraordinary variability is particularly informative. Thus, for example, the frequencies of the different C types (Figure 8) clearly distinguish European Caucasoid, African Black and Japanese populations, providing a sort of fingerprint for each population. The definition is statistical, based on the collection of frequencies and does not, of course, provide a method for assigning unequivocally any individual to one or other of these populations. That may only be done with any assurance when there is a gene that is known to be present only in one population, and not the other. This happens for some disease genes such as that connected with the sickle cell trait, whose origin is in West Africa and is connected with resistance to malaria.

The overall information provided by a set of frequencies, such as those for the HLA types, can be combined using appropriate statistical techniques to give a phylogenetic tree of the relationship between the different populations studied. This is based first on using measures of similarity between the populations based on the patterns of gene frequencies, and then on algorithms which group together those populations that are most similar and leads to the construction of a tree based on objective criteria of similarity between the populations. Such trees are generally quite consistent with conventional classifications of the relationships between different human population groups based on historical, linguistic, physical and archaeological evidence (Figure 9). Thus, African, Caucasoid and
Figure 9. A phylogenetic tree showing the relationship among eighty different populations, constructed using the ‘neighbour joining’ method from frequencies of HLA-A and B genes. The length of the arms is a measurement of the ‘genetic’ distance separating the populations. (Courtesy of T. Gojobori from Tsuji et al. 1992.)
Oriental populations are clearly separated. Intriguingly, the Mongolian populations are placed at the boundary between the Orientals and the Caucasoids. In general there is an overall remarkably good correlation, as pointed out by Luca Cavalli-Sforza and his colleagues, between these genetically defined population relationships and linguistic and cultural classifications. This clearly does not necessarily extend to finer classifications of linguistic differences, though it is, for example, notable that the Czechs, Poles, Hungarians and Slovaks form a closely knit group. Clearly, there is a point at which the distinction between geographical on the one hand and linguistic and other cultural determinants of populations on the other is hard to disentangle. The British and Cornish, including no doubt a number of Celts, sit fairly and squarely in the midst of the Northern European as do, still, the North Americans and the Australians.

Although A1 is characteristic of Caucasian populations as a whole, the combination A1 B8 is characteristic particularly of Northern European populations, and seems to have its highest frequency in those who share, in general, a Celtic origin (Figure 10). These data have been assembled and analysed by Julia Bodmer, based on extensive tissue typing results obtained throughout the United Kingdom from laboratories doing typing

![Figure 10](image)

**Figure 10.** The frequency of the HLA-A1, B8 haplotype in Welsh, Scottish, Irish and English populations. The numbers above the columns are the frequency × 1000. (Courtesy of Julia Bodmer.)
for transplant matching. The high frequency in Wales and Scotland is quite striking and suggests that this combination may indeed, to some extent be a marker of Celtic origin, at least for the British Isles.

The DNA Revolution

The traditional genetic markers of population studies have been the blood groups, and more recently systems such as the HLA tissue typing system, that essentially detect chemical differences whose inheritance is clearly determined. The discovery, however, of the structure of DNA, the chemical substance of the genes in which is encoded the true language of life, together with our ability to manipulate DNA and read its message, has revolutionized the range of individual differences that can be identified. Now, and in the future, most studies of individual differences will investigate them at the level of the DNA, namely the gene itself.
Chromosomes are made of DNA (Figure 11) which is a double helical molecule, as shown by Watson and Crick, whose two strands are made up of sequences of four letters A, T, G and C. There is a single defined sequence because of the pairing rule which determines that opposite an A there is a T, and opposite a G a C. The genetic language is, therefore, a four letter language and it is the sequence of these letters that determines our inherited make-up. A gene is a particular segment of this sequence connected with a particular function. The total number of DNA letters in the human instruction set is three billion, namely three thousand million, \((3 \times 10^9)\) or approximately a hundred million per chromosome.

Many techniques now exist for detecting differences at the DNA level. One example, the first to be used in population and family studies and still widely used, depends on the use of enzymes which will cut the DNA only where there is a particular sequence. If individuals vary with respect to such a sequence, then in one case an enzyme may cut the DNA at a particular place while for another individual it will not. This results in patterns of bands that differ between individuals as a function of their DNA sequence at a particular position (Figure 12). Such differences can

![Family 79](image)

**Figure 12.** Family DNA typing. Circles are females, squares males, filled in symbols are individuals affected with the disease 'polyposis'. The pattern of bands below each individual is their DNA type for the particular difference being studied. The top band is the same in all individuals while the bottom two, labelled A1 and A2 can vary. In this family individuals are either A1 and A2 or A2 only, the latter associating with the disease. (From Bodmer et al. 1987, Nature, 328: 614–16.)
be used to track diseases through families and are the basis on which new disease genes are now being found. These are the same techniques that should eventually identify the specific genes that determine facial features and other attributes by which we recognize each other, and which have so far formed the conventional basis for racial classification. The range of differences at the DNA level that can be studied in this way is essentially unlimited. This means that it is possible to characterize populations with greater and greater precision, provided enough DNA differences are studied in enough individuals from each population. The possible applications of this powerful approach to the study of populations and their interrelationships are only just beginning to be exploited.

European Populations and the Spread of Agriculture

A phylogenetic tree (courtesy of Cavalli-Sforza and colleagues) of twenty-six European populations, based on eighty-eight genes, using the range of
markers that is now available gives a more precise definition of their inter-relationships than was possible before (Figure 13). The tree is reasonably consistent with geographical and linguistic inter-relationships, for example, placing the Dutch, Danish and English into a closely knit group adjacent to the Swiss, Germans, Belgians and Austrians, and clearly indicating as has been known for some time, that the Lapps and Sardinians are significant outliers from the remaining European populations. The Scottish and Irish Celtic representatives are, in this tree, clearly separated from the English indicating again, as in the case of the HLA-A1 B8 combination, the distinctiveness derived from the Celtic ancestry of these populations.

The spread of agriculture across Europe, from its centre of origin in the Middle East starting some ten thousand years ago, has, as pointed out by Luca Cavalli-Sforza and his colleagues, been a major determinant of

Figure 14. The spread of agriculture across Europe. The lines represent the time of first arrival in thousands of years before the present. The data are based on carbon dating of the earliest farming culture remains known in each region. (From Bodmer and Cavalli-Sforza 1976 page 545.)
Figure 15. Synthetic map of gene frequencies across Europe. This represents the first principle component calculated from the same data used for figure 13. (Courtesy L. L. Cavalli-Sforza from P. Menozzi, A. Piazza and L. Cavalli-Sforza, History and Geography of Human Genes, Princeton University Press, In Press.)

the distribution of genetic differences across Europe (Figure 14). Thus, patterns of genetic difference gradients across Europe in many cases parallel the spread of agriculture. A statistical technique known as Principle Components Analysis can analyse similar patterns of frequency changes using data on many populations and many genes together. The first principle component is essentially a linear combination of those genes whose frequencies give the most similar pattern of variation between the different populations. The results of such an analysis by Cavalli-Sforza et al. using essentially the same populations as were used for construction of the phylogenetic tree, and a similar set of genetic markers of which a third are from the HLA System, are shown in Figure 15. The gradient of frequency change, spreading out from the Middle East and paralleling the spread of agriculture, is quite striking. Cavalli-Sforza and his colleagues have emphasized that this is exactly what would be expected if farmers had migrated from the centre of agriculture, slowly at each point inter-marrying with the local population. Though the number of migrants might be small, before agriculture population densities were low and so a relatively small number of migrants could have had a significant impact on
the genetic constitution of a population after inter-marriage. Agriculture would then have enabled a considerable increase in population size, fuelling the migrational advance and continuing the dilution of genes whose frequencies differed between populations at the centre of origin of agriculture and those at its northern periphery. These data show a remarkable agreement between the genetic analysis of population differences and the archaeological evidence concerning the spread of agriculture.

One of the most striking features of European populations is the predominance of a light skin, a trait that clearly seems to have spread from Northern Europe southwards. This is associated with lighter pigmentation of the hair and of the eyes (Figure 16). There seems to be no doubt that this combination of traits must have had its origin somewhere in the north, perhaps in Northern Scandinavia and gradually spread southwards.

Skin colour is largely genetically determined. A light skin provides an advantage in an environment with low sunlight, since it enhances the formation of vitamin D, whose insufficiency leads to rickets which must be a considerable disadvantage for both survival and reproduction. Conversely a light skin is a grave disadvantage in a tropical climate, not only
because of damage to the skin by the ultra-violet light from the sun, but also, perhaps, because of a risk of excess production of vitamin D. It seems possible that a mutation for light skinned pigmentation was initially selected for because of its advantage, in terms of natural selection, in the extreme North of Europe, with its deficiency of sun-light. The clear disadvantage of this skin colour in the tropics can be seen from the very high incidence of skin cancer of whites in tropical areas, such as in Australia, a risk which is particularly high for red heads with the least skin pigmentation, and so a particular disadvantage for such Celts when they migrated to Australia. A further factor in the eventual spread of light-skin colour, together with lighter hair, both brown and red, and lightly pigmented eyes, either blue or green, is likely to have been sexual selection, namely preferential mating with those who had these unusual and no doubt striking features.

This southward spread of characteristics is paralleled by some other gene frequency patterns across European populations. There may thus have been a collection of genes whose frequency was high in Northern Europe and which has spread southwards. These genes no doubt parallel those determining skin, hair and eye colour, but the latter spread more extensively presumably because of sexual selection. Their association with the other genetic markers would then gradually be diluted out.

The Origin of the Celts of the British Isles

How do these data on the patterns of migration across Europe and the genetic differences between the northern and southern populations relate to the origin of the Celtic populations in the British Isles? In my undoubtedly superficial reading of the Celtic populations, I have been struck by two apparent paradoxes.

The first is the contrast between some of the classic descriptions of Celts, such as that of Queen Boudicca, or of the Gauls, as tall, handsome and blond, and the conventional picture of the Welsh and the Central European Celts, as short and black, or dark haired.

The second is the emphasis on the description of the earlier Iron age Celtic settlements in Central Europe and the possible relationship of this to Celtic invasions of the British Isles, around 600 or 700 BC and even later incursions as a result of the Roman conquest of Gaul. These events seem much too recent, and too long after the spread of agriculture, to have had such a major influence on population distributions in the British Isles. For a significant influence from such incursions, the ratio of the invaders to the invaded, in terms of eventual settlements and genetic contributions, would
have to be very considerable. This could happen if the invaders displaced the indigenous population, as no doubt happened to some extent, for example, with the Anglo-Saxon invasions of the east coast of Britain and, of course, has happened much more obviously with the colonisation of North America. However, it seems possible, and indeed likely, that the relatively late Celtic invasions from Continental Europe, while they may have had a considerable cultural impact and, for example, influenced the spread of the Celtic language, may nevertheless have contributed comparatively little to genetic change. This is even more the case for the Roman occupation of Britain, which seems to have left almost no genetic trace.

Agriculture came to the British Isles about five or six thousand years ago. Before that the Isles were probably occupied by a small number of hunter-gatherers (perhaps no more than ten to fifty thousand) who were descended from the late Palaeolithic hunters who occupied North Western Europe. These presumably were the populations that are the origin of light skin, red or brown hair and those genes that are most frequent in the north and have a decreasing frequency gradient to the south. The farmers who were at the tail end of the spread of agriculture from the Middle East that came to the British Isles around five and a half thousand years ago, will have brought their admixture of genes as it developed from the pattern of migration that followed the spread of agriculture across Europe. This presumably may be the main origin of the dark hair and small stature. No doubt these migrating farmers, who may be the closest relatives of the Celts of Central Europe, fancied the blond, blue-eyed women of the north. The Celts of the British Isles therefore could really be a distinctive mixture of the original hunter-gatherer populations of the far north of Europe, with their blond features and high HLA-A1 B8 and relatively low type A frequencies, and the darker skinned and dark hair coloured farmers migrating from the middle of Europe outwards, bringing their new agricultural skills with them that led to a major population expansion. It is perhaps this expanded mixed population which represents the true Celts of the British Isles and which is thus quite different from the Celtic populations of Central Europe. The common culture and language may have come in part with the farmers, but also with later invasions being a source of cultural diffusion, but in that case without genetic change. The Celts as defined in this way probably once occupied most of the British Isles, as is suggested by the distribution of megalithic monuments. Their concentration in the fringes, namely in Wales, Ireland, Scotland and Cornwall, was presumably due to later invasions, down even to the Norman conquest. Preliminary genetic market studies suggest that Cornwall has retained less of its Celtic identity than Wales (especially North Wales),
probably because of greater admixture with the English. Scotland and Ireland on the other hand were more contaminated by the Vikings.

The Handbook of Man

The complete instruction set for a human being, containing some three thousand million letters, is equivalent to two hundred thousand pages of text each containing some fifteen thousand letters worth of information. This is about equivalent to ten copies of the new complete Oxford English Dictionary. The Handbook of Man will eventually be completely deciphered. That is the goal of the Human Genome Project. Fortunately, for the scientists at least, for reasons that are not understood only about 10 per cent of the DNA sequences actually are used and informative, and so the information that matters is about equivalent to the content of one copy of the complete Oxford English Dictionary. This, however, exists in an almost infinite variety of versions because of the genetic differences between individuals. It is these differences that, as I have emphasized, provide the basis for a genetic characterization of populations and their history. As the complete handbook is deciphered, new opportunities for the more precise genetic definition of population differences, especially with respect to those facial and other outward features by which we traditionally characterize populations, will continually emerge. These will provide an extraordinary opportunity for combining historical, cultural and archaeological analysis of populations with the increasingly precise genetic descriptions.

I hope you will excuse me if I have made some outrageous suggestions because of my limited knowledge of the Celtic populations, their culture and their history. On the other hand I do hope that I may have persuaded you of the enormous power of combining humanistic studies of human populations with the scientific investigation of genetic differences.

References