A Y Chromosome Census of the British Isles

Cristian Capelli,[1,8] Nicola Redhead,[1]
Julia K. Abernethy,[1] Fiona Gratrix,[1]
James F. Wilson,[1] Torolf Moen,[3] Tor Hervig,[4]
Martin Richards,[5] Michael P.H. Stumpf,[1,9]
Peter A. Underhill,[6] Paul Bradshaw,[7] Alom Shaha,[7]
Mark G. Thomas,[1,2] Neal Bradman,[1,2]
and David B. Goldstein[1]
[1]Department of Biology
[2]The Centre for Genetic Anthropology
University College London, UK
Gower Street
London WC1E 6BT
United Kingdom
[3]Trondheim University Hospital
N-7006 Trondheim
Norway
[4]Haukeland University Hospital Blood Bank
N-5021 Haukeland
Denmark
[5]Department of Chemical and Biological Sciences
University of Huddersfield
Huddersfield HD1 3DH
United Kingdom
[6]Department of Genetics
Stanford University
Stanford, California 94305-5120
[7]BBC Archaeology
London
United Kingdom
[8]Istituto di Medicina Legale
Università Cattolica di Roma
Roma I-00168
Italy
[9]Department of Zoology
Oxford University
Oxford OX1 3PS
United Kingdom

Summary

The degree of population replacement in the British Isles associated with cultural changes has been extensively debated [1–3]. Recent work has demonstrated that comparisons of genetic variation in the British Isles and on the European Continent can illuminate specific demographic processes in the history of the British Isles. For example, Wilson et al. [4] used the similarity of Basque and Celtic Y chromosomes to argue for genetic continuity from the Upper Palaeolithic to the present in the paternal history of these populations (see also [5]). Differences in the Y chromosome composition of these groups also suggested genetic signatures of Norwegian influence in the Orkney Islands north of the Scottish mainland, an important center of Viking activities between 800 and 1300 A.D.

*Correspondence: d.goldstein@ucl.ac.uk

[6]. More recently, Weale et al. [7] argued for substantial Anglo-Saxon male migration into central England based on the analysis of eight British sample sets collected on an east-west transect across England and Wales. To provide a more complete assessment of the paternal genetic history of the British Isles, we have compared the Y chromosome composition of multiple geographically distant British sample sets with collections from Norway (two sites), Denmark, and Germany and with collections from central Ireland, representing, respectively, the putative invading and the indigenous populations. By analyzing 1772 Y chromosomes from 25 predominantly small urban locations, we found that different parts of the British Isles have sharply different paternal histories; the degree of population replacement and genetic continuity shows systematic variation across the sampled areas.

Results and Discussion

To represent the indigenous population of the British Isles, we have selected a site in central Ireland that has had no known history of contact with Anglo-Saxon or Viking invaders (Castlerea, see Figure 1). Given the demonstrated similarity of Celtic and Basque Y chromosomes [4, 5] (p = 0.6, using haplogroups), these sample sets were combined [8, 9] to provide a representation of the Y chromosomes of the indigenous population of the British Isles. Norwegian invaders were represented by two sites in western Norway (Bergen and Trondheim), Danes were represented by a general Danish collection, and Anglo-Saxons were represented by samples from their historical homeland in Schleswig-Holstein (North Germany). Linguistic and historical investigations seem to suggest that internal migrations were minor and have not unduly blurred the genetic landscape of North Germany and Denmark in the last 1500 years [10]. We also note that some historians view the Anglo-Saxons themselves as Germanic invaders from what is now North Germany/Denmark. Population differentiation between the continental and indigenous British Isles groups was assessed by using an analog of Fisher's exact test calculated by using haplogroup (hg) frequencies, as implemented by the Arlequin software package [11]. There was no significant difference between the Trondheim and Bergen samples or between the Danish and North German samples (p = 0.8), while the Norwegians were different from the other northern European samples (p < 0.05). We therefore clustered these source populations into two continental groups, referred to from now on as the Norway and German/Danish sample sets. Note that the similarity of the Danish and North German Y chromosomes means that, at the hg resolution, we cannot distinguish the genetic contributions to the British Isles of the two component groups. All continental populations, however, show significant differences from the indigenous group (p < 0.01), and Norway can be distinguished, though to a lesser degree, from the German/Danish sample (p <

1.Shetland
2.Orkney
3.Durness
4.Western Isles
5.Stonehaven
6.Pitlochry
7.Oban
8.Morpeth
9.Penrith
10.Isle of Man
11.York
12.Southwell
13.Uttoxeter
14.Llanidloes
15.Llangefni
16.Rush (Dublin)
17.Castlerea
18.Norfolk
19.Haverfordwest
20.Chippenham
21.Faversham
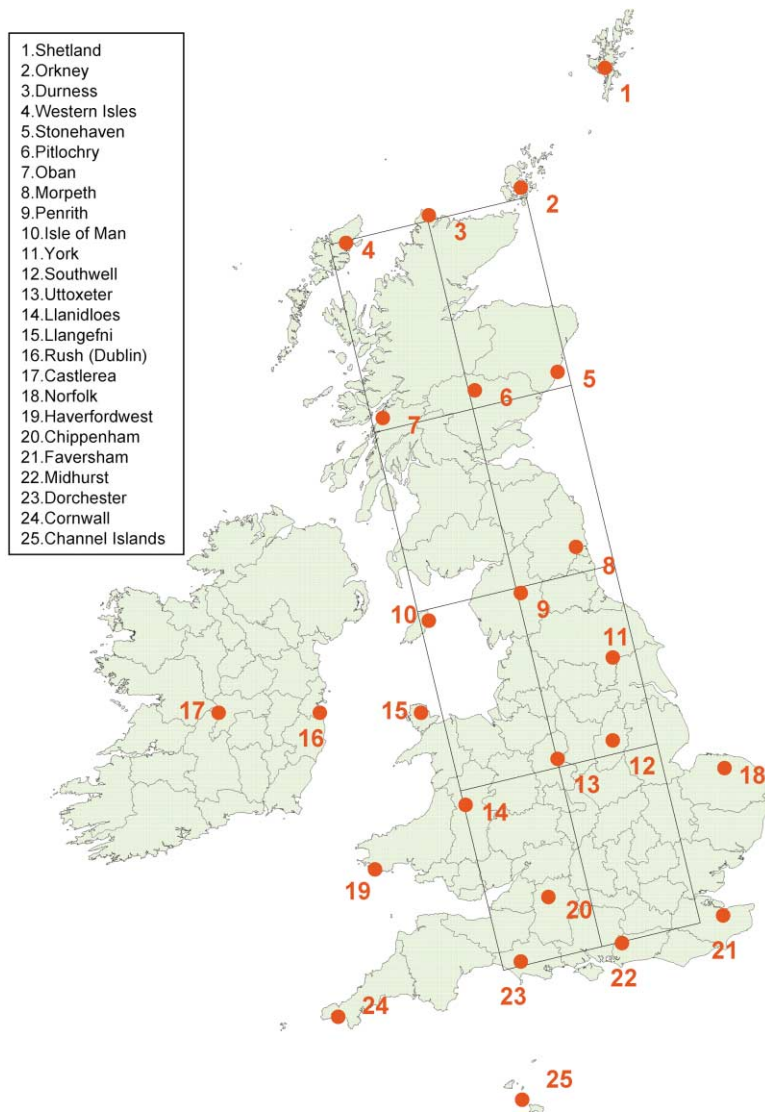22.Midhurst
23.Dorchester
24.Cornwall
25.Channel Islands

Figure 1. British Isles Sampling Locations Map

The location of the sampled small, urban areas and the 3 × 5 grid of collection points are shown. For each grid point, we selected the closest town within a 20-mile radius. Only towns with 5–20,000 inhabitants were chosen. Individuals were, with the exception of one location, then selected if their paternal grandfather's birthplace was within a 20-mile radius of the selected center. Midhurst samples were collected up to 40 miles from the respective grid point. When the grid point was at sea, the nearest point on the coast was used (Morpeth and Stonehaven). We also added additional points to cover important geographic regions not covered by the grid (Shetland, York, Norfolk, Haverfordwest, Llangefni, Chippenham, Cornwall, Channel Islands) and included two Irish samples, Castlerea and Rush (North of Dublin). The total number of points sampled in the British Isles was 25.

0.05). Sampling in the British Isles was mainly undertaken to conform to a systematic 3 × 5 grid (Figure 1).

We genotyped six Y-linked microsatellites (identifying haplotypes) [12] and 11 unique event polymorphisms (UEPs, identifying hgs, as defined in Figure 2) known to be polymorphic in Europe [13, 14]. The most frequent haplogroups observed were those defined by M173, M170, and M17 mutations (Table 1) (Hgs R1xR1a1, IxI1b2, and R1a1, referred to as hgs 1, 2, and 3 in [4]). For most analyses, we subdivided chromosomes within
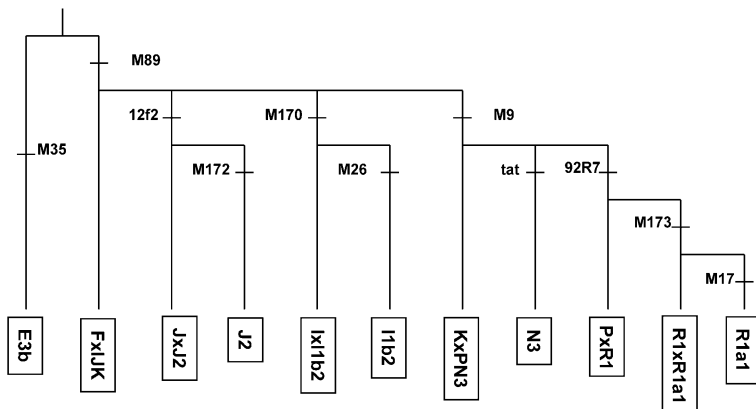
Figure 2. Y Chromosome Genealogy

Y chromosome genealogy of the UEPs typed is shown. Nomenclature is as suggested by the Y chromosome consortium [18]. For simplicity, only the last derived mutation is indicated in the text.

Table 1. Y Chromosome Haplogroup Frequencies in the Different Populations

Sample	E3b	FxIJK	JxJ2	J2	IxI1b2	2.47+1	I1b2	KxPN3	N3	PxR1	R1xR1a1	AMH+1	R1a1	3.65+1	n
Shetland					0.05	0.05					0.17	0.49	0.06	0.17	63
Orkney					0.07	0.07	0.01			0.02	0.23	0.41	0.07	0.12	121
Durness					0.02	0.12					0.47	0.33	0.02	0.04	51
Western Isles					0.18	0.07					0.17	0.49	0.03	0.06	88
Stonehaven		0.02			0.02	0.11					0.34	0.45	0.05		44
Pitlochry				0.07	0.10						0.24	0.56		0.02	41
Oban		0.02			0.05	0.02					0.26	0.60	0.02	0.02	42
Morpeth		0.02	0.01	0.03	0.12	0.06					0.21	0.52	0.02	0.01	95
Penrith	0.03	0.01		0.02	0.08	0.10					0.16	0.52	0.02	0.06	90
Isle of Man	0.02				0.08	0.08					0.15	0.55	0.08	0.05	62
York	0.04	0.02			0.15	0.17					0.20	0.37	0.02	0.02	46
Southwell	0.06			0.06	0.14	0.04					0.20	0.44	0.04	0.01	70
Uttoxeter	0.04	0.01		0.04	0.08	0.10					0.26	0.45		0.02	84
Llanidloes	0.05	0.04		0.02	0.07	0.12					0.19	0.47	0.04		57
Llangefni	0.04			0.01	0.04			0.01			0.21	0.68	0.01		80
Rush					0.08		0.03				0.45	0.41	0.01	0.03	76
Castlerea					0.07		0.02				0.37	0.53			43
Norfolk	0.03			0.02	0.17	0.14					0.23	0.37	0.02	0.02	121
Haverfordwest	0.03				0.02		0.02				0.27	0.64	0.02		59
Chippenham		0.02		0.04	0.06	0.14	0.02				0.16	0.49	0.06	0.02	51
Faversham	0.04			0.05	0.04	0.07					0.27	0.49	0.02	0.02	55
Midhurst	0.01	0.01	0.01	0.04	0.09	0.06	0.03				0.20	0.54	0.01		80
Dorchester	0.04	0.01	0.01	0.03	0.07	0.07					0.26	0.47	0.04		73
Cornwall					0.02	0.04	0.08				0.25	0.54	0.06	0.02	52
Channel Islands	0.04	0.02	0.01	0.01	0.10	0.11	0.03				0.27	0.39	0.02	0.01	128
Basques	0.02	0.02			0.02		0.05				0.29	0.60			42
Germany/Denmark	0.03	0.02		0.03	0.19	0.20			0.02		0.13	0.26	0.08	0.04	190
Norway		a		a	0.13	0.15			0.01	0.04	0.08	0.22	0.12	0.22	201

R1xR1a1, IxI1b2, and R1a1 have been indicated by subtracting AMH+1, 2.47+1, and 3.65+1, respectively. a, frequency of 0.005.

these groups by using one-step neighbor clusters of the haplotypes AMH, 2.47, and 3.65 [4], indicated as AMH+1, 2.47+1, and 3.65+1, respectively.

In Principal Component (PC) plots summarizing variation in Y chromosome frequencies (Figure 3), all the British populations (excluding Orkney and Shetland) skewed toward the right of axis one; this reflects the relatively high frequencies of AMH+1 in these populations. The Basques, the most extreme on this axis, clustered with samples from central Ireland (Castlerea) and Wales (Haverfordwest and Llangefni). It is interesting to note that Scottish mainland sites appear generally between English ones and these "indigenous" populations. The Norwegian and the German/Danish samples are separated on axis two.

To aid interpretation of this plot, we simulated admixed populations by drawing varying proportions of individuals at random from each of the source populations and plotting them on PC plots that also included the source populations themselves (Figure 4). The simulated populations show that the position of a population on the first two axes provides a sensitive indicator of the degree of continental input into an indigenous background, with Norwegian input moving populations strictly along axis 1 and German/Danish input moving populations at an angle through both axes. Inspection of Figure 3, in light of these simulations, shows that Orkney and Shetland have significant Norwegian input and little to no German/Danish input, that the English and Scottish sites all have German/Danish influence, and that the Western Isles and Isle of Man have German/Danish influence, presumably due to English immigra-

tion. In addition to these relatively clear patterns, the PC plots also provide a suggestion of more subtle differences. For example, there is a group of populations that appears shifted to the left from the main angle of German/Danish influence, and this is consistent with some degree of Norwegian input. It is not surprising that the Western Isles and Isle of Man are in this group, but the inclusion of Penrith is of particular interest given the Scandinavian influence on dialect in this region [15]. Similarly, Rush appears to be shifted slightly toward the Norwegian pole on PC1, but there is no shift toward the German/Danish position. In addition, the mainland Scots are somewhat closer to the indigenous type than any English sets, except Cornwall. The sites with the highest degree of German/Danish input are York and Norfolk, followed by Southwell and Llanidloes. All of these except Llanidloes are historically in regions where the Danes are known to have had a significant presence. The peculiar position of Llanidloes might reflect recent migration in the past two centuries [1]. The remaining samples are closer to the indigenous group; for these populations, this finding suggests a lower demographic impact by North European populations. This can also be seen in the frequency of AMH+1, which is always above 33% in British populations but remains below 26% in the continental source populations; these data are consistent with the presence of some indigenous component in all British regions.

Admixture proportions were also evaluated by using a likelihood approach implemented by the program Lea [16]. This quantitative analysis was consistent with the visual pattern shown by PC investigation, and it also
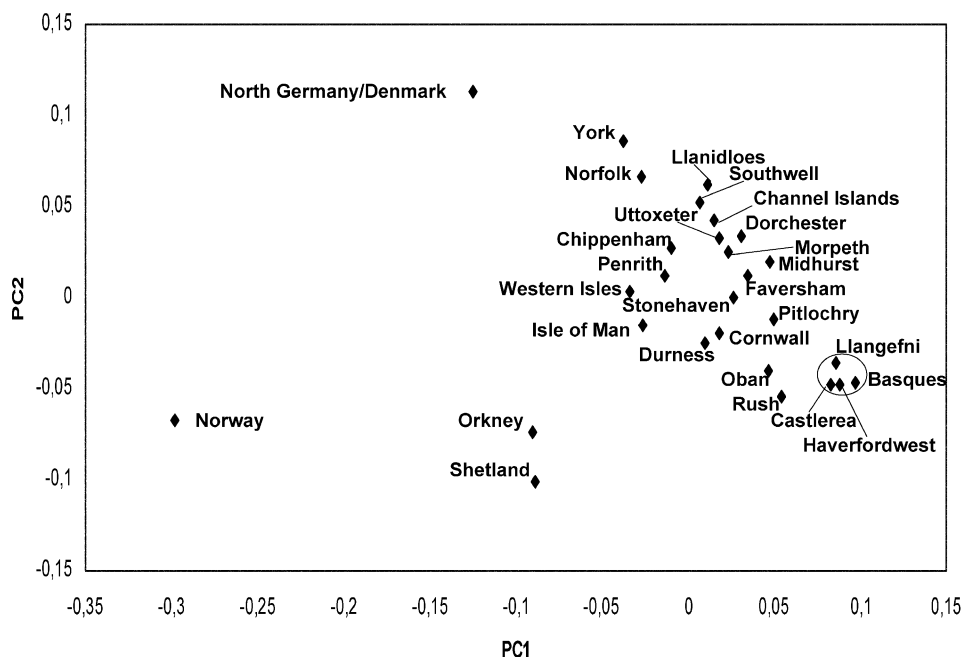
Figure 3. Principal Components Plot

A plot of the first and second principal components of the Y chromosome haplogroup frequencies of the populations shown in Table 1. The first two components of the Principal Components analysis of Y chromosome frequencies explain almost 60% of the total variation. The loadings with the greatest magnitude for the first axis are for AMH+1 and 3.65+1 (+0.152 and −0.241), while 2.47+1 and 3.65+1 have the greatest impact on the second axis (+0.128, −0.131).

provides significant evidence that there has not been complete population replacement anywhere in the British Isles (see Table S1 in the Supplemental Data available with this article online).

The apportionment of genetic variation was inferred with AMOVA, as implemented by the Arlequin package [11]. Comparison of the different small towns sampled indicates that the vast majority of the diversity present was within populations (96.35%), with only 3.65% across populations. The subdivision of the samples into Celtic (Ireland, Wales, and mainland Scotland) versus the rest of the populations showed a distribution across the two groups of 3.65% of the total variation; the exclusion of Llanidloes and Durness, which clearly show evidence of continental input, increased this value to 6.16% (Figure 3 and Table S1). Considering the indigenous/nonindigenous clustering system (Castlerea, Haverfordwest, and Llangefni versus the rest), a value of 7.48% was calculated, one of the highest values obtained, among multiple alternative clustering systems (not shown). Thus, the indigenous/nonindigenous distinction appears to be the most important factor influencing geographic patterns of Y chromosome variation in the British Isles.

In summary, our results show that Norwegian invaders heavily influenced the northern area of the British Isles, but this group had limited impact through most of mainland Scotland (except the extreme north). Instead, mainland Scotland was more influenced by the German/Danish input. Despite their well-known activities in the Irish Sea, Norwegian input in adjoining areas is modest. Some is indicated in the Isle of Man, and a smaller amount is indicated in Ireland. Perhaps the most surprising conclu-

sion is the limited continental input in southern England, which appears to be predominantly indigenous and, by some analyses, no more influenced by the continental invaders than is mainland Scotland (Figure 3 and Table S1). It is interesting to note that the areas in southern England were, historically, mostly occupied by the Anglo-Saxons, while the activities of the Danish Vikings were mainly in eastern England [1]. The results seem to suggest that in England the Danes had a greater demographic impact than the Anglo-Saxons. An alternative explanation would be that the invaders in the two areas were genetically different and that we cannot see this difference reflected in the current inhabitants of the Continental areas corresponding to Anglo-Saxon and Danish homelands. This would seem to be a difficult distinction to make, and it should be emphasized that our analyses assume that we have correctly identified the source populations. If, for example, the real continental invaders had a composition more similar to the indigenous British than our candidate sample set, our results would systematically underestimate the continental input. Similarly, any Continental input into our Castlerea sample would bias our inferences, but the very similar composition of the Basque and Castlerea samples suggests that this has been minimal. With regard to source populations, we note that Weale et al. [7] recently used Friesland as an Anglo-Saxon representative source population and suggested a substantial replacement of pre-Anglo-Saxon paternal lineages in central England. We therefore compared Frisians to our North German/Danish sample and found that the two sets are not significantly different from each other (p =
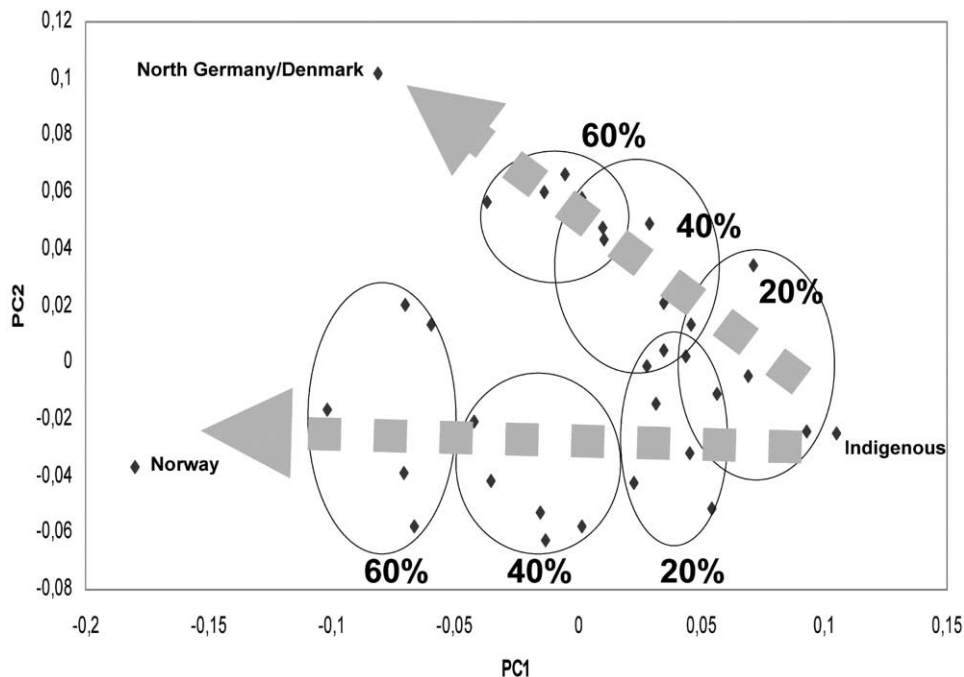
Figure 4. Principle Components Plot of Simulated Populations

A Principle Components plot of simulated populations of n = 50 comprising, at admixtures of, respectively, 20%, 40%, and 60%, the indigenous British and the German/Danish and Norwegian sets. The circles group simulated populations with the same continental proportions. The arrows indicate the directions along which the simulated population tends to move according to the relative proportion of Continental input.

0.3, data not shown). When included in the PC analysis, the Frisians were more "Continental" than any of the British samples, although they were somewhat closer to the British ones than the North German/Denmark sample. For example, the part of mainland Britain that has the most Continental input is Central England, but even here the AMH+1 frequency, not below 44% (Southwell), is higher than the 35% observed in the Frisians. These results demonstrate that even with the choice of Frisians as a source for the Anglo-Saxons, there is a clear indication of a continuing indigenous component in the English paternal genetic makeup. We also note that our analysis includes representatives of the Danish Vikings, which were not available in the Weale et al. study. Consideration of Danish Viking input is important because their activities on the British eastern coast are well documented [1]. Our evaluation of the Danish and Anglo-Saxon source populations, however, shows that the contributions of these groups are unlikely to be distinguishable by using the resolution available in our analyses. Whatever level of replacement took place in England, it could have been due to "Anglo-Saxons," Danes, or a combination of both groups.

Conclusions

The detailed sampling scheme used here identified other previously unknown regional patterns in the degree of continental input. For example, the Central-Eastern part of England experienced the most continental introgression. In addition, our inclusion of samples from Wales additional to those of Weale et al. [7] indicates that the transition between England and Wales is somewhat

gradual, which was not visible in the samples analyzed in the Weale et al. study

Most studies in human evolution and genetic history have used samples from very few locations, often near major metropolitan areas. Here, we show that detailed samples from multiple small, urban areas with a geographically structured sampling design reveal patterns that could not be detected with typical sampling schemes. For example, analyses of multiple sets have confirmed higher continental input in central England and the northernmost samples (Durness, on the north coast of Scotland and the Scottish Isles) and a lower level of continental introgression in southern England and Lowland Scotland. In addition, multiple sample sets revealed heterogeneity in Wales.

Iberian, French, and Central-Northern Italian populations have been shown to have similar Y chromosome compositions, presumably reflecting their common heritage in the European Palaeolithic [14]; Wilson et al. [4] noted that AMH+1 haplotypes at high frequency are associated with the European Palaeolithic. Here, we note that another haplogroup (I1b2) is found almost exclusively in British populations that have experienced little or no continental genetic input (Tables 1 and S1). Intriguingly, earlier studies have shown that it is present in the Iberian Peninsula at low frequencies (0%–5.4%) and in Sardinia at a significant percentage (35.1%) [9, 14]. This group might be another constituent of the European Palaeolithic.

Finally, we note that forensic analyses based on the Y chromosome generally assume homogeneity of Y chromosome haplotypes throughout most of Europe

[17]. Our fine-scale investigation of Y chromosome variation demonstrates appreciable frequency differences of Y chromosome haplotypes over relatively short geographic distances. Haplotype 12 13 11 16 25 11 (hg R1a1) (number of repeats, loci as follows: DYS388, 393, 392, 19, 390, 391) is present at frequencies around 5% in Shetland and Orkney, while it is almost completely absent from the other collections. Similarly, haplotype 14 13 11 14 22 10 (hg lxl1b2) was recorded at 6%–7% in the Central-East English samples, but it was absent from Irish, Welsh, and Scottish populations.

Experimental Procedures

Microsatellite and UEP Analysis

Y chromosome microsatellites DYS388, 393, 392, 19, 390, and 391 analysis was performed by following the protocols described [12]. UEP analysis was based on a PCR-RFLP approach. Protocols will be published elsewhere and are available from C.C. upon request. Briefly, the DNA region containing the chosen polymorphic nucleotides was PCR amplified and then screened by using appropriate restriction endonucleases. Digested PCR products were loaded on a 377 ABI automated sequencer updated to 96 lanes, and alleles were called according to fragment size.

Data Analysis

Principal Components analysis was performed by using the POPSTR software (H. Harpending, personal communication). The apportionment of genetic variation and Fisher's Exact Test analog were inferred by using the Arlequin package [11].

Supplemental Data

Supplemental Data including Table S1 are available at http://images.cellpress.com/supmat/supmatin.htm.

References

1. Davies, N. (1999). The Isles: A History (London: Macmillan).
2. Graham-Campbell, J., and Batey, C. (1998). Vikings in Scotland (Edinburgh: Edinburgh University Press).
3. Richards, J.D. (2000).Viking Age England (Stroud: Tempus).
4. Wilson, J.F., Weiss, D.A., Richards, M., Thomas, M., Bradman, N., and Goldstein, D.B. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc. Natl. Acad. Sci. USA *98*, 5078–5083.
5. Hill, E.W., Jobling, M.A., and Bradley, D.G. (2000). Y-chromosome variation and Irish origins. Nature *404*, 351–352.
6. Berry, R.J., and Firth, H.N. (1986). The People of Orkney (Kirkwall: Orkney Press).
7. Weale, M.E., Weiss, D.A., Jager, R.F., Bradman, N., and Thomas, M. (2002). Y chromosome evidence for Anglo-Saxon mass migration. Mol. Biol. Evol. *19*, 1008–1021.
8. Bosch, E., Calafell, F., Santos, F.R., Perez-Lazaun, A., Comas, D., Benchemsi, N., Tyler-Smith, C., and Bertranpetit, J. (1999). Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. Am. J. Hum. Genet. *65*, 1623–1638.
9. Bosch, E., Calafell, F., Comas, D., Oefner, P.J., Underhill, P.U., and Bertranpetit, J. (2001). High-resolution analysis of human Y-chromosome variation shows a sharp discontinuity and limited gene flow between northwestern Africa and the Iberian Peninsula. Am. J. Hum. Genet. *68*, 1019–1029.
10. Forster, P. (1995). Einwanderungsgeschichte Norddeutschlands (Immigration history of North Germany). In North-Western European Language Evolution, V.F. Faltings, A.G.H. Walker, and O. Wilts, eds. (Odense, Denmark: University of Odense), pp. 141–163.
11. Schneider, S., Kueffer, J.-M., Roessli, D., and Excoffier, L. (1997). ARLEQUIN, A Population Genetic Data Analysis Programme. (Geneva: Genetics and Biometery Laboratory, University of Geneva).
12. Thomas, M.G., Bradman, N., and Flinn, H.M. (1999). High throughput analysis of 10 microsatellite and 11 diallelic polymorphisms on the human Y-chromosome. Hum. Genet. *105*, 577–581.
13. Rosser, Z.H., Zerjal, T., Hurles, M.E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., et al. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. Am. J. Hum. Genet. *67*, 1526–1543.
14. Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., et al. (2000). The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. Science *290*, 1155–1159.
15. Reaney, P.H. (1927). A Grammar of the Dialect of Penrith (Cumberland): Descriptive and Historical, with Specimens and Glossary (Manchester, UK: The University Press).
16. Chikhi, L., Bruford, M.W., and Beaumont, M.A. (2001). Estimation of admixture proportions: a likelihood-based approach using Markov chain Monte Carlo. Genetics *158*, 1347–1362.
17. Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., et al. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci. Int. *118*, 106–113.
18. Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res. *12*, 339–348.