Genetic evidence for different male and female roles during cultural transition in the British Isles

James F. Wilson[†‡], Deborah A. Weiss[§], Martin Richards[†], Mark G. Thomas[¶], Neil Bradman[¶], and David B. Goldstein[†‖]

Gatsby Lab at Department of Biology, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom; Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, United Kingdom; §Department of Anthropology, University of California, Davis, CA 95616; and Department of Genetics and Biometry, University College London, Gower Street, London WC1E 6BT, United Kingdom

Communicated by Henry C. Harpending, University of Utah, Salt Lake City, UT, January 23, 2001 (received for review July 7, 2000)

Human history is punctuated by periods of rapid cultural change. Although archeologists have developed a range of models to describe cultural transitions, in most real examples we do not know whether the processes involved the movement of people or the movement of culture only. With a series of relatively well defined cultural transitions, the British Isles present an ideal opportunity to assess the demographic context of cultural change. Important transitions after the first Paleolithic settlements include the Neolithic, the development of Iron Age cultures, and various historical invasions from continental Europe. Here we show that patterns of Y-chromosome variation indicate that the Neolithic and Iron Age transitions in the British Isles occurred without large-scale male movements. The more recent invasions from Scandinavia, on the other hand, appear to have left a significant paternal genetic legacy. In contrast, patterns of mtDNA and X-chromosome variation indicate that one or more of these pre-Anglo-Saxon cultural revolutions had a major effect on the maternal genetic heritage of

26—AG-GGGTACC+G; and hg 28—AG-GGGTGCC+G. A tree presenting the genealogical relationships of these hgs (except hg 28, which branches from hg 26) is presented in ref. 11.

mtDNA Hgs. Haplotypes were assigned to hgs according to the West Eurasian mtDNA genealogy (12). hg assignment proceeded by using the following algorithm (all numbering is according to ref. 13 minus 16,000 in the control region for brevity): 069T 126C 223C assigned to hg J (note in all but four cases 069 information was available); 126C 223C 294T assigned to T; 129A 223T 391A assigned to I (391 information was available); 223T 292T assigned to W; 189C 223T 278T assigned to X; 223C 224C 311C assigned to K; 223C 249C and either 189C or 327T assigned to U1; 129C 223C assigned to U2 (051G, if information available); 223C 343G assigned to U3; 223C 356C assigned to U4; 223C 270T assigned to U5; 172C 219G 223C assigned to U6; 223C 318T assigned to U7; 223C 298C assigned to V; 067T 223C assigned to HV1 (067 information usually available); 126C 223C 362C assigned to preHV; 145A 176G 223T assigned to N1b; 223T 278T 390A assigned to L2; and 187T 189C 223T 278T 311C assigned to L1. For sequences not matching any of those above, the algorithm used was the following: if 223T, test for +10397 *AluI* (where + indicates restriction site presence and − indicates absence) for M; −10871 *MnlI* and −10397 *AluI* for L1, L2, or L3; if 223C, test for −7025 *AluI* for H; −14766 *MseI*, +7025 *AluI*, −4577 *NlaIII* for HV*; +12308 *HinfI* for U*, otherwise assign to R*. The first hypervariable section (HVS-1) sequences were checked also for matches with common East Asian hg motifs. Recurrent mutations may cause ambiguities by eliminating part of a diagnostic motif or recreating it in another part of the tree. In many cases, the presence of substitutions defining subclades within the major hgs allowed sequences to be assigned even when reversion had occurred at an hg motif site. In the case of hybrid motifs, PCR-restriction fragment length polymorphism (RFLP) testing was used to assign hg (14). Particularly in the Welsh and Irish data, HVS-1 sequences matching a unique haplotype in an RFLP-defined hg were assigned to that hg.

Analysis. Exact tests and analyses of molecular variance were calculated by using ARLEQUIN (15). Principal components analyses were performed on hg and allele frequencies by using POPSTR (H. Harpending, personal communication). Population structure was assessed by using the model-based clustering method implemented in STRUCTURE (16). The admixture model was used with a burn-in of 50,000 steps and a run length of 10^6 steps. All loci within 2 centimorgans of another locus were excluded from the STRUCTURE analysis, leaving 23 loci.

Results and Discussion

Genetic History of Orkney. When the Norsemen invaded (about A.D. 800), Orkney was populated by the Picts, little-understood pre-Anglo-Saxon inhabitants. Orkney remained a Norse colony while an increasing number of Scottish settlers arrived in the islands, which were pledged to Scotland in 1468 (17). As the place-names of Orkney are almost entirely Old Norse in origin (18) and a Nordic language replaced the earlier tongue, linguists have assumed that the Viking invaders completely replaced the native population (19). Modern archeological interpretations, however, suggest continuities in both artifacts and lifestyle, which are more compatible with considerable integration between native Picts and incoming Norsemen (20, 21). To investigate whether Orkney's Viking heritage is genetic as well as cultural, we sampled 71 adult males claiming at least three unrelated paternal generations in Orkney, and all with surnames found on the islands before 1700 (22). For comparison, we used analogous criteria to sample 78, 88, and 94 individuals from Norway, Anglesey (North Wales), and West Friesland (The Netherlands), respectively. Data on 146 Irish males with Irish Gaelic surnames also were included (23).

The Irish and Welsh are not significantly differentiated from each other at the hg level [= 0.16 (24)] and will hereafter be called "Celtic." However, Celtic, Frisian, Norwegian, and Orcadian Y chromosomes are all highly differentiated at the hg level (< 0.0001) (Fig. 1 and Table 1). The Orkney sample seems interme-

(AMH). In each of the Basque, Welsh, and Irish populations, a total of 89–90% of the chromosomes are in hg 1, which contains the M173-defined Eu18 hg in Semino . (34), with the majority of the remainder in hg 2. The Turkish sample, however, is much more

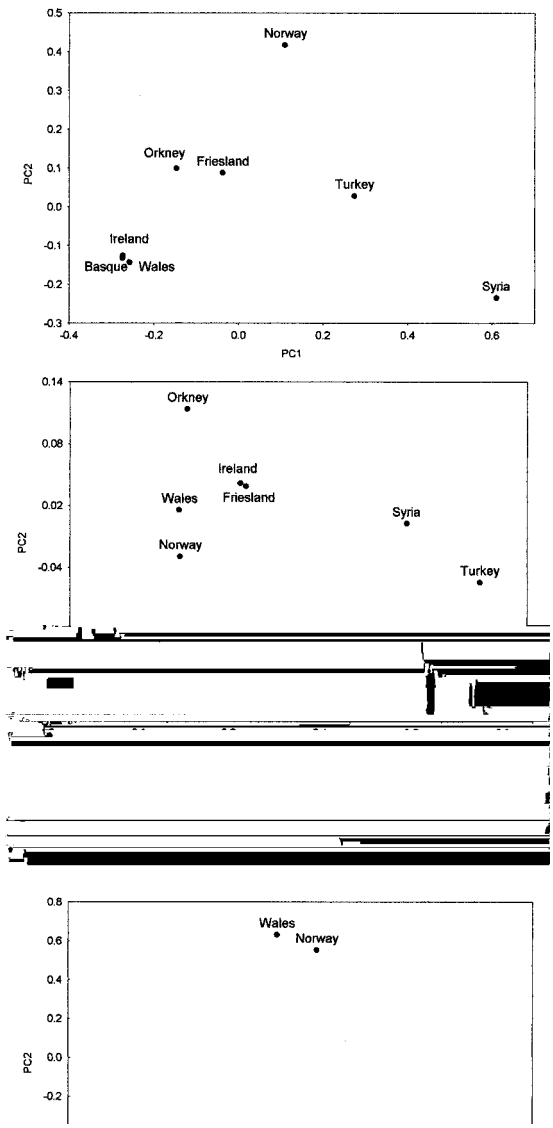

Fig. 2. P t t a PC Y (*Top*), t DNA (*Middle*), a X I q »t (*Bottom*) a, , i -/ t b t I . A, Y (I , (J t f» (I E a t DNA (I : H, V, J, T, I, W, X, U3, U4, U5, a K. I t Y- q a t t I t 'a q I , a I b-/PC1 I 65% a t q , a I b-/PC2 I 28%, (a I t t DNA a X I q »t q a t (a 54% a 15%, a 46% a 33%, t f–, -/

156 Basques (33, 48), 101 Irish (33), 218 Turks (33), and 69 Syrians (33). Slowly evolving coding-region variants and control-region sites are used to assign mtDNAs into genealogical clades or hgs, whereas more quickly evolving control-region sites define haplotypes within hgs. The hg distributions in all of the European populations are very similar (9, 49). Turkey and Syria, however, are distinct with much lower frequencies of the most common European hg (H) and large proportions of hgs not present or extremely rare in the European samples. The lack of structure is also evident at the haplotype level of resolution; analysis of molecular variance apportions 99% of the variance in our European populations between individuals within populations, regardless of the grouping scheme.

Principal Components (PC) Analysis. PC analyses were performed on both Y chromosome and mtDNA hg frequencies (Fig. 2). In each case, the first PC (explaining 65% and 54% of the variation, respectively) depicts a general East–West population gradient; a pattern usually interpreted as indicating the Neolithic component (32, 50). In line with this interpretation, the poles of the first PC of both systems are defined on the one hand by the Basques, and on the other by Turkey and Syria. As may be expected, in the Y-chromosome plot, the Celtic-speaking populations fall extremely close to the Basques, and Orkney falls midway between the Atlantic cluster and Norway. This pattern is in sharp contrast to that for mtDNA, in which the Celtic-speaking populations are closer to the center of the plot, indicating that they have undergone more female-mediated gene flow from other European populations than the Basques have. Thus, at least one of the cultural transitions in the British Isles since the Upper Paleolithic must have involved a demic component on the female side. The similarity of the non-Basque European populations means that there is no power to apportion the Orcadian maternal heritage into Scandinavian and pre-Anglo-Saxon British components by using the available mtDNA data.

X-Chromosome Microsatellites. To assess which of the two uniparentally inherited genetic systems more closely reflects the history of the genome more widely and to check that the lack of differentiation among the British and non-Basque European populations is not caused by a lack of resolution in the mtDNA data, we analyzed microsatellites on the X chromosome. Although having far less genealogical information at each genetic locus than is available for completely linked systems such as mtDNA and the Y chromosome, multilocus genotypes are known to provide a sensitive test of population structure (16, 51). Thirty-four dinucleotide markers located across the length of the X chromosome were genotyped in the Basques, Norwegians, Welsh, and Turks. Population structure was assessed by using a model-based clustering approach implemented in the STRUCTURE program (16). Briefly, the model assumes K populations, each characterized by a set of allele frequencies at each locus, and individuals are assigned to these populations on the basis of their genotypes. We estimated Pr($|K$), where is the data, for $K = \{1, 2, 3, 4\}$. By using Bayes' theorem and assuming a uniform prior on K between 1 and 4, we can then approximate the posterior distribution, Pr($K|$). For the Basque, Welsh, Norwegian, and Turkish data, all of the posterior probability is on $K =$ 1, i.e., there is no detectable genetic structure.

However, when we performed a PC analysis on the allele frequencies at these 34 X-linked microsatellites, we observed a pattern essentially identical to that seen for mtDNA (Fig. 2). Once more, the Basques and Turks occupy opposite poles of PC1 and the Welsh and Norwegians fall in the center of the plot. Despite there being no statistical support for genetic structuring in the X-microsatellite data considered on their own, the similarity of the patterns observed across different genetic systems provides robust evidence that the Basques are differentiated from the other European populations, specifically in having a lower input from the Near East.

Female-mediated gene flow between the Celtic-speaking populations and other North European populations has thus homogenized the variation, not only for mtDNA but also for other parts of the genome affected by female migration. There are two extreme scenarios that could account for the sharp differences observed between the genetic systems that are and that are not (Y chromosome) affected by female movement (mtDNA, X chromosome, and the Y chromosome, respectively). First, the pre-Anglo-Saxon British source populations may have been different from the current European population for the Y chromosome but less so for other regions of the genome. This explanation is inconsistent with the position of the Basques, however, which is distinctive for both the Y chromosome and the systems affected by female migration. The second explanation is that the European Paleolithic populations were originally dis-

tinct from the current European population for both the Y chromosome and other parts of the genome, but this distinctiveness was eroded subsequently by female movements between the Celtic-speaking and non-Basque European populations. In other words, at least one of the Neolithic or Iron Age cultural transitions in the British Isles involved some female immigration.

Population parameters such as estimates of divergence times inferred from one-locus systems always have a high variance, because information is only incorporated from one realization of the evolutionary process. Certain evolutionary questions, however, are less subject to this source of variation and can be addressed profitably with only a single genetic locus. For example, identification of related lineages in different populations could be taken as secure evidence of some kind of connection between the populations such as gene flow or common ancestry, even though genetic drift at a single locus would make it impossible to estimate accurately parameters reflecting the quantitative relationship (e.g., migration rate or population-separation time). Despite these problems, in cases where female migrations have homogenized the variation in other parts of the genome, the Y chromosome may be the only signal of certain historical relationships.

In summary, we have identified markers of paternal Scandinavian influence in the British Isles that suggest the Viking settlement of Orkney involved substantial genetic as well as cultural replacement. Accepting the widely held view that the Basques are representative of pre-Neolithic European Y chromosomes (32), we have also shown that Neolithic, Iron Age, and subsequent cultural revolutions had little effect on the paternal genetic landscape of the Celtic-speaking populations (there has been continuity from the Upper Paleolithic to the present). However, comparison with mtDNA and X-linked microsatellites reveals that at least one of these cultural revolutions had a major effect on the maternal genetic heritage of the Celtic-speaking populations.

Note Added in Proof. Basque, Welsh, Norwegian, and Orcadian hg 1 chromosomes also were genotyped at $DYS194_{469}$ and 25/25, 72/75, 18/20, and 45/46, respectively, carried the derived A allele [i.e., were hg 1L in the nomenclature of Hammer . . (52)].

We thank the subjects who provided samples; M. A. Strøksnes, S. Jones, and R. Jager for collecting samples; J. Betranpetit for the Basque DNA; and E. Hill, D. Bradley, C. Tyler-Smith, L. L. Cavalli-Sforza, M. Jobling, T. Zerjal, and F. Calafell for access to unpublished results and useful discussions.

1. Hawkes, C. (1931) *A . .* **5,** 60–97.
2. Renfrew, C. (1987) *A . . J .* (Penguin, London).
3. Cunliffe, B. (1997) *. A . C .* (Oxford Univ. Press, Oxford).
4. Jobling, M. A. & Tyler-Smith, C. (1995) *. G . .* **11,** 449–456.
5. Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F. R., Schiefenhovel, W., Fretwell, N., Jobling, M. A., Harihara, S., . . (1997) *A . J. H . G . .* **60,** 1174–1183.
6. Karafet, T. M., Zegura, S. L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., . . (1999) *A . J. H . G . .* **64,** 817–831.
7. Thomas, M. G., Skorecki, K., Ben-Ami, H., Parfitt, T., Bradman, N. & Goldstein, D. B. (1998) *N . . (L . .)* **394,** 138–140.
8. Thomas, M. G., Bradman, N. & Flinn, H. M. (1999) *H . G . .* **105,** 577–581.
9. Richards, M., Corte-Real, H., Forster, P., Macaulay, V., Wilkinson-Herbots, H., Demaine, A., Papiha, S., Hedges, R., Bandelt, H. J. & Sykes, B. (1996) *A . J. H . G . .* **59,** 185–203.
10. Wilson, J. F. & Goldstein, D. B. (2000) *A . J. H . G . .* **67,** 926–935.
11. Zerjal, T., Pandya, A., Santos, F. R., Adhikari, R., Tarazona, E., Kayser, M., Evgrafov, O., Singh, L., Thangaraj, K., Destro-Bisol, G., . . (1999) in *G . D . : A . . H . . . G . .*, eds. Papiha, S. S. & Deka, R. (Plenum, New York), pp. 91–102.
12. Macaulay, V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne-Tamir, B., Sykes, B. & Torroni, A. (1999) *A . J. H . G . .* **64,** 232–249.
13. Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., . . (1981) *N . . (L . .)* **290,** 457–465.
14. Torroni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996) *G . .* **144,** 1835–1850.
15. Schneider, S., Kueffer, J.-M., Roessli, D. & Excoffier, L. (1997) ARLEQUIN, A Population Genetic Data Analysis Program (Genetics and Biometry Laboratory, Univ. of Geneva, Switzerland).
16. Pritchard, J. K., Stephens, M. & Donnelly, P. (2000) *G . .* **155,** 945–959.
17. Thomson, W. P. L. (1986) in . . . . , eds. Berry, R. J. & Firth, H. N. (Orkney Press, Kirkwall), pp. 209–224.
18. Lamb, G. (1993) . . . . . (Byrgisey, Kirkwall).
19. Barnes, M. P. (1998) . . *N . J . . . . .* (Shetland Times, Lerwick).