mtDNA and the Islands of the North Atlantic: Estimating the Proportions of Norse and Gaelic Ancestry

Agnar Helgason,[1,3] Eileen Hickey,[2], Sara Goodacre,[2] Vidar Bosnes,[4] Kári Stefánsson,[3]
Ryk Ward,

though Norn, the Norse dialect of Orkney and Shetland, survived until the 19th century). Figure 1 shows the geographic location of the North Atlantic islands, the sailing routes of the Vikings, and the areas of Norse influence in the North Atlantic region.

A number of previous studies have attempted to shed light on the genetic affinities of the North Atlantic islanders using classical allozyme genetic markers, but their results have been difficult to interpret. Most studies have focused on the Icelanders, with the aim of calculating the contributions to admixture of the Norse and Gaelic ancestral populations. Estimates have varied considerably—from a 93%–98% Gaelic contribution (Thompson 1973) to an 86% Norse contribution (Wijsman 1984). More recent analyses of mtDNA and Y-chromosome variation in the Icelanders suggests that a majority of the female settlers may have originated from the British Isles, whereas ~80% of male settlers were Scandinavian (Helgason et al. 2000*b,* 2000*c*). Fewer studies have dealt with the other island populations in the North Atlantic. An analysis of classical allozyme markers in a Western Isles population from Lewis found that allele frequencies showed substantial differences from neighboring European populations (Clegg et al. 1985). Natural selection or genetic drift in the Lewis

gene pool or gene flow among the other European populations were suggested as possible causes of these differences. In similar studies of Orkney and Shetland, Roberts reported (1985, 1990) that both island populations diverged considerably in allele frequencies from neighboring populations. Although not ruling out selection or drift as potential causes for this divergence, Roberts concluded that the islanders of Orkney and Shetland most likely represented remnants of an aboriginal gene pool that had changed on the British mainland because of later population movements. None of these studies of allozyme variation in the Scottish Islands reported estimates of Norse admixture.

In this study, we examined mtDNA control-region sequences in the North Atlantic island populations of Orkney, the Western Isles, the Isle of Skye, and Iceland, and we compared them to those observed in the rest of the British Isles, Scandinavia, and other regions of Europe. The primary aims were to assess the relative magnitude of diversity and levels of Gaelic and Scandinavian admixture in the mtDNA pools of the North Atlantic island populations and individuals from the northwest coast of Scotland. mtDNA lineages sampled from contemporary populations provide us with direct links to matrilineal ancestors from the Viking age, and thus en-
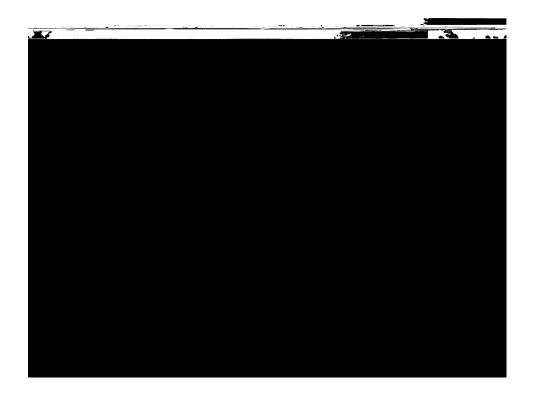

Figure 1   Map of the North Atlantic region. The blackened area represents regions where Norse cultural and linguistic dominance was complete during the Viking period. The dark speckled areas in the British Isles provide an indication of the core areas of Viking exploits, on the basis of archaeological sites, place names of Norse origin, and raided monasteries and towns (Bjarnason et al. 1973; Graham-Campbell and Batey 1998; Corráin 1999; Keynes 1999). The arrows show some of the main sailing routes of the Vikings in the North Atlantic.

ables us to examine the extent to which Scandinavian females were involved in Norse settlements on the North Atlantic islands. Many historians believe that the Norse expansion of the Viking Age was primarily a male enterprise (Clover 1988). If this were the case, one would not expect to find close links between mtDNA lineages found in the North Atlantic island populations and Scandinavia. On the basis of available historical, archaeological, and linguistic information (Graham-Campbell and Batey 1998; Corráin 1999; Davies 1999), we would expect that the largest proportion of mtDNA lineages inherited from Norse matrilineal ancestors would be found (in descending order of magnitude) in Iceland, Orkney, the Western Isles, the Isle of Skye, and the coastal region of northwest Scotland. To achieve these aims, we sequenced the mtDNA hypervariable segment 1 (HVS1) from 1,664 individuals from the Scottish islands, the Scottish mainland, England, and Norway. These new data were added to an existing data set of 3,444 Eurasian HVS1 sequences, for a detailed study of mtDNA variation in the North Atlantic region.

Material and Methods

Population Samples

DNA from 891 individuals from all regions of mainland Scotland, 181 from the Western Isles, 49 from the Isle of Skye, and 142 individuals of English matrilineal descent (representing most regions of England) was extracted from blood collected at Blood Transfusion Service donor sessions throughout Scotland. Information about the birth place of the maternal grandmother was sought from each individual. Similarly, DNA from 323 Norwegians was extracted from blood collected at donor sessions at Ullevål hospital in Oslo. Again, the birth place of the maternal grandmother was recorded for each individual, showing the samples to be representative of all geographic areas of Norway. In all cases individuals gave informed consent. Samples from 78 Orkney Islanders were kindly supplied by Dr. J. Bodmer (see Bodmer et al. 1996). The data produced for this study were deposited in GenBank and are available on request from the corresponding author.

The comparative data set from Europe consisted of 3,444 mtDNA HVS1 sequences from the following populations and sources: Iceland (394 from Helgason et al. 2000*b*, 20 from the Mitochondrial DNA Concordance, 14 from Richards et al. 1996, and 39 from Sajantila et al. 1995), Ireland (23 from the Mitochondrial DNA Concordance and 105 from Richards et al. 2000), Orkney (74 from the Mitochondrial DNA Concordance), the Western Isles (16 from the Mitochondrial DNA Concordance), England and Wales (160 from Richards et al. 1996, 29 from the Mitochondrial DNA Concordance,

and 97 from Piercy et al. 1993), Finland (74 from Kittles

or on a 377 DNA sequencer (Perkin-Elmer). In most cases, sequences were obtained between sites 16010 and 16400. The mtDNA site numbers referred to in this study are those of Anderson et al. (1981). To maximize the number of sequences available for analysis from the North Atlantic region, all analyses were restricted to the 235 nucleotides between positions 16090 and 16324. Recent studies have indicated that European mtDNA pools contain an extensive array of different mtDNA
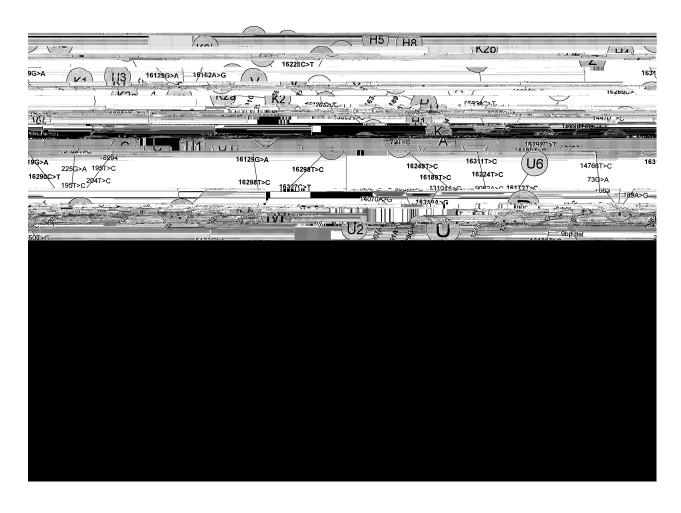
Figure 2 Schematic phylogenetic representation of mtDNA lineage clusters found in European populations, reconstructed on the basis of information obtained from Torroni et al. (1996), Richards et al. (1998), Macaulay et al. (1999), Quintana-Murci et al. (1999), and Helgason et al. (2000*b*

means of a multidimensional scaling (MDS) analysis, using the SPSS software package.

Distances

Using the full sequences between sites 16090 and 16324, we estimated the mutational divergence of the North Atlantic mtDNA pools from other European populations using the index . The index is defined as the average number of substitutions between the sequences of one population and the closest founder sequences observed in another population (Forster et al. 1996), and it effectively summarizes the overlap between one mtDNA pool and a potential source mtDNA pool. Unlike an analysis of molecular variance (AMOVA) distance, which summarizes the average mutational distance between all pairs of sequences from two populations (Excoffier et al. 1992), is insensitive to the fact that the divergence between European populations, as measured in mutations at the mtDNA locus, is small relative to the overall mutational time-depth of the European mtDNA phylogeny (see Richards et al. 1998; Simoni et al. 2000a). Thus, an AMOVA analysis shows that <2% of the variance in mutational divergence between all pairs of European mtDNA HVS1 sequences are accounted for by their distribution among different populations (Helgason et al. 2000b).

Admixture Method

To estimate the level of Scandinavian ancestry in the island populations of the North Atlantic and the coastal population of northwest Scotland, we employ a heuristic approach to estimate the admixture proportion that best fits the observed lineage distribution in the admixed and parental populations (see Helgason et al. 2000c). This estimator, designated m , is obtained as follows. Given a prior probability of admixture, η, the probability that a randomly chosen lineage observed in the admixed sample is derived from the first source population is given by $\eta p_1 / \{\eta p_1 + (1 - \eta)p_2\}$, where p_1 and p_2 are the frequencies of this lineage in the net47lage popula.250(20/F14 1 Tf07vy626 T1(the0 Tc[(quenci)-370(t20.61a0)]TJ-370(t20.61a0)]TJ

Table 1

Summary Statistics for HVS1 Sequences from European Populations

Population	N	K	S	GD	π	k	

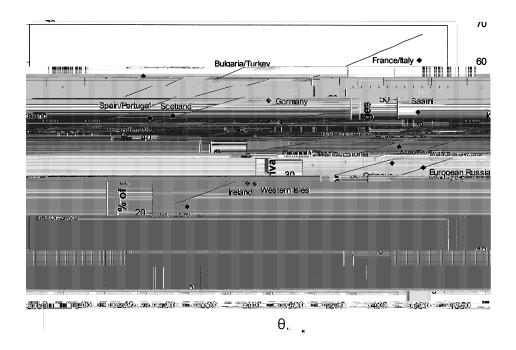
70
60
40
30
20
Bulgaria/Turkey
France/Italy
Spain/Portugal
Scotland
Germany
Saami
Iceland
European Russia
Orkney
Ireland
Western Isles
Isle of Skye
% of private lineages
θk

Figure 3 Scatterplot of $_k$ values and the percentage of private lineages. The least-squares regression line, where $r = 0.77$, is shown. The curved lines show the 95% confidence region around the regression line.

Isle of Skye and the northwest coast of Scotland and is greatest for the Icelanders. The least difference between Scandinavians and Gaels is observed for Iceland and then, in descending order, for Orkney, the Western Isles, the northwest coast of Scotland, and the Isle of Skye. Interpreted as a rough indicator of Norse admixture, the relative differences between distances to Gaels and Scandinavians accord with historical evidence of the differential impact of Norse settlements on each of the North Atlantic island populations. However, according to these results, it appears that the Gaelic contribution to the Icelandic mtDNA pool may have been at least as large as that from Scandinavia.

The small distances to the Germans in all five cases are surprising, as there are no known accounts of recent female gene flow from Germany into the North Atlantic region. This may be accounted for by Germany's central position in Europe and by the fact that many ancient population movements into the British Isles and Scandinavia originated from or passed through this territory

Table 2

Pattern of Lineage Sharing between North Atlantic Islands and Source Populations

		LINEAGES SHARED WITH (%)			
POPULATION	K[a]	Gaels Only[b]	Scandinavians Only[b]	Both	Neither
Iceland	114	11.4 (.62)	7.0 (.38)	26.3	55.3
Orkney	68	16.2 (.65)	8.8 (.35)	38.2	36.8
Western Isles	79	16.5 (.81)	3.8 (.19)	35.4	44.3
Isle of Skye	23	30.4 (.78)	8.7 (.22)	39.1	21.7
NW Scottish coast	91	24.2 (.88)	3.3 (.12)	39.6	33.0
Scottish Islands	138	18.8 (.74)	6.5 (.26)	27.5	47.1
North Atlantic islands	214	14.0 (.67)	7.0 (.33)	21.0	57.9

[a] K = number of distinct lineages.
[b] The number in parentheses represents the proportion of lineages shared exclusively with either the Gaels or Scandinavians out of the total number of exclusively shared lineages.

strength and significance of the apparent geographic structure observed in figure 5. Geographic distances were calculated as geodesic distances, using the following coordinates for each population: Austria/Swiss (15°12′E, 48°42′N), European Russia (35°42′E, 57°0′N), Finland/Estonia (25°12′E, 60°6′N), France/Italy (7°24′E, 44°6′N), Germany (10°12′E, 51°0′N), Iceland (18°24′W, 64°42′N), Ireland (7°42′W, 53°23′N), Orkney (2°54′W, 59°17′N), Scandinavia (11°18′E, 59°30′N), Scotland (4°18′W, 56°30′N), Bulgaria/Turkey (29°48′E, 39°17′N), Spain/Portugal (30°W, 39°42′N), England/Wales (2°6′W, 52°42′N), and the Western Isles/ Isle of Skye (7°6′W, 57°17′N). The product moment correlation between genetic and geographic distances for all the groups in figure 5 was $r = 0.717$, and, of 10,000 random permutations of the distance matrices, none yielded values $\geqslant 0.717$. When the North Atlantic island populations are omitted from the matrices used in the Mantel test, we obtain $r = 0.713$, with the same high degree of significance.

Estimates of Admixture

In this section, we apply the heuristic approach to estimation of the relative contributions of the Gaelic and Scandinavian source populations to the mtDNA pools of the island and coastal populations of the North Atlantic. Table 4 shows the estimated ancestral proportions for each of the five admixed populations, along with 95% confidence intervals.

The estimate of Scandinavian ancestry ranges from 11.5% in the Western Isles to 37.5% in the Icelanders. For the Scottish island and coastal populations, these findings are consistent with historical, archaeological, and linguistic evidence of Norse settlements in the North Atlantic region. Of the Scottish populations, Orkney evidently has the closest matrilineal links with Scandinavia. The Western Isles, the Isle of Skye, and the coastal pop-

ulation of northwest Scotland all exhibit similarly low levels of Scandinavian mtDNA ancestry. In contrast to expectations based on historical records, the Icelanders have a similar proportion of Scandinavian mtDNA ancestry to that of the Orkney islanders, indicating that the majority of Icelandic matrilines originated from the British Isles.

Discussion

This study of mtDNA variation in the North Atlantic demonstrates the utility of uniparental loci in contributing to the understanding of the origins of human populations not only in broad terms over millennia but also in regional studies well within the historical period. The analysis has revealed important details of the relationships and demographic histories of the North Atlantic island populations. The islanders of Skye are clearly identified as the least diverse of the North Atlantic island populations. The populations of Iceland, Orkney, and the Western Isles exhibit greater levels of genetic diversity, and judging from ϑ_k values, had similar female effective population sizes ( )
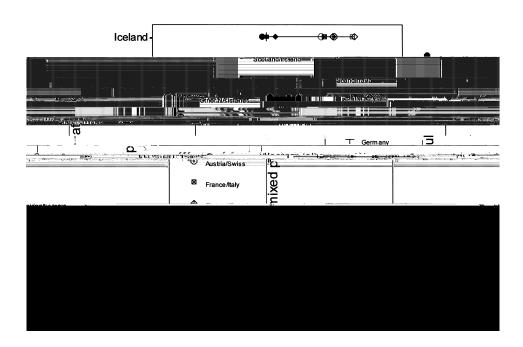
Figure 4 distances between the mtDNA pools of five North Atlantic populations and those of potential European source populations. distances to the Saami were excluded to maximize clarity in the representation of distances to the other populations. In all cases, the distance to the Saami was ~1.5.

in earlier centuries. Thus, the reported $_k$ values do seem to reflect recent historical female effective population sizes for all three islands.

On the basis of an expected positive association between $_k$ values and the proportion of private lineages across populations, we observed a relative excess of private lineages in the Icelanders compared with the Scottish island populations. Although this may be explained partially by greater sampling saturation in Iceland, it is also likely to reflect the geographic isolation of the Icelanders, which will have hindered gene flow to and from the island. The relative lack of private lineages in the Scottish islands suggests higher levels of gene flow and accords with the recorded depopulation of the last 200 years, which would have brought many private island lineages to the Scottish mainland.

As is the case with most European populations, there is some overlap between the mtDNA pools of the Scandinavians and Gaels. Of a total of 416 lineages found in these two populations, 73 (17.5%) were shared. Between 35% and 40% of lineages found in the Scottish islands and 26% of those found in Iceland are shared with both the Scandinavian and Gaelic populations. However, there are differences between these mtDNA pools that can be exploited to shed light on the genetic history of the North Atlantic island populations. The inhabitants of the Scottish islands share two to seven times more of their lineages exclusively with Gaels than

they do with Scandinavians. This difference is smaller for the Icelanders, but nonetheless indicates a closer link to Gaels.

The pattern of lineage sharing is also informative about interrelationships among the North Atlantic island mtDNA pools. Of the lineages whose distribution is restricted to the North Atlantic islands, only two are not private lineages (one belongs to haplogroup V and has the substitution motif 16124C 16298C 16362C; the other belongs to subcluster T2, with the motif 16093C 16126C 16153A 16294T). In both cases, the lineages are shared between Icelanders and the Western Islanders. Intriguingly, the medieval record of the settlement of Iceland (*The Book of Settlements* 1972) indicates that the Western Isles were frequently the place of departure for settlement voyages to Iceland and that a number of indigenous Western Islanders accompanied such voyages. In all, the Western Isles are mentioned 22 times in *The Book of Settlements,* Orkney is mentioned 7 times, the Faroe Islands are mentioned 3 times, and Shetland is mentioned 2 times. Although anecdotal, this apparent link between Iceland and the Western Isles suggests that mtDNA lineages can be used to identify recent migration contributions from closely related populations.

The different estimates of Scandinavian ancestry for mtDNA lineages in Orkney, the Western Isles, the Isle of Skye, and the northwest coastal region of Scotland
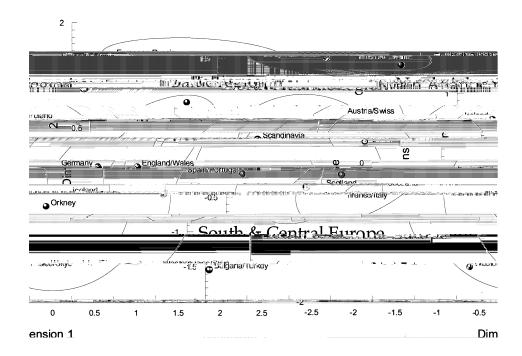
Figure 5   Multidimensional scaling plot of genetic distances on the basis of haplogroup frequencies. The fourteen dimensions of the genetic distance matrix were reduced to two dimensions, which account for 85% of the genetic variation defined by the original distance matrix.

are consistent with the intensity of Norse activities. Although Norse influence extended to all these places, the Earldom of Orkney (established by the Norwegian king in 900 A.D.) was the political and strategic hub of Norse activities in the North Atlantic region. The sheer number of Norse archaeological sites in Orkney and historical accounts in medieval texts, such as *Orkneyinga Saga,* testify to the thorough Norse occupation of this island group. Clover (1988) has pointed out that, like many human range expansions, the movement of Norse people during the Viking period was male dominated. Thus, it is recorded that initial Viking activities in the British Isles involved raiding parties of Norse men (Davies 1999; Sawyer 1999). After these early raids and for >50 years before Iceland was discovered and colonized, many of the same Viking men settled and intermarried with existing populations in Shetland, Orkney, the Western Isles, the Isle of Man, and coastal regions of Ireland, Scotland, and northern England (Jones 1984; Davies 1999; Sawyer 1999). In Orkney, it would have been natural for whole families to move from Scandinavia—those of the Earl, his kinsfolk, and their retainers. However, most of the Norse who raided, traded, and settled elsewhere in the British Isles were young and

Table 3

Haplogroup and Subcluster Frequencies for European Populations

	FREQUENCY FOR POPULATION (%)														
HAPLOGROUP	Austria/ Switzerland (N = 187)	European Russia (N = 215)	Finland/ Estonia (N = 202)	France/ Italy (N = 248)	Germany (N = 527)	Iceland (N = 467)	Ireland (N = 128)	Orkney (N = 152)	Scandinavia (N = 645)	Scotland (N = 891)	Bulgaria/ Turkey (N = 102)	Spain/ Portugal (N = 352)	England/ Wales (N = 429)	Western Isles/Isle of Skye (N = 246)	Saami (N = 176)
A	0	0	0	0	0	0	0	0	.16	0	0	.85	.23	.41	0
B	0	0	0	0	0	0	0	0	0	.11	0	0	0	0	0
C	0	1.86	.50	.40	.19	.43	0	0	0	0	1.96	1.14	0	0	0
D	0	1.86	0	.40	.38	0	0	0	.16	0	4.90	.28	0	0	5.11
H	47.06	33.49	36.63	45.16	38.33	28.27	41.41	40.79	39.69	38.38	31.37	50.28	40.79	27.24	1.70
H1	3.74	5.58	2.97	2.82	4.36	8.35	2.34	6.58	3.41	3.03	1.96	2.56	4.43	1.22	1.14
H3	.53	0	0	.40	1.14	.43	1.56	0	.62	.56	0	0	.93	1.22	0
H4	1.60	0	1.49	1.61	1.52	9.64	2.34	.66	2.33	1.23	3.92	3.41	3.03	3.25	0
H5	1.07	0	0	2.82	.76	.43	0	0	.62	.11	0	1.99	1.63	0	0
H8	0	3.26	2.97	.81	2.85	.43	0	2.63	1.86	2.36	.98	.28	1.40	1.63	2.84
I	2.14	1.40	2.48	.81	2.28	4.71	2.34	3.29	1.86	4.38	1.96	.57	3.03	6.50	0
J	5.35	6.51	4.95	2.42	6.83	6.85									
J	5.3573 1 Tf0..2857 TD3/F4 1 Tf6									0 N42(2.34)-3434247(347(4.38-346c2)-3447(9.64)-3e7(9)-71535(0)-56877(9)-7153					

I6.51   87(1.(.62)57-49)-3447(U5a1.662)-14

8   0   U5b1.6618157(9)453(2279*[(37(53([(49)-237957-8)Tj/.)4837(1)84(000)Tj2797()-8)Tj/F14 142247(Tj7(69298)Tj/.)4837(1)84(000)Tj25441438)Tj/F14 1442497(Tj-844328480(0)-850(.).)4837(1)48U95((

Table 4

Scandinavian and Gaelic Mitochondrial Ancestry in the North Atlantic Islanders and the Coastal Population of Northwest Scotland

	m (95% CI) FOR ANCESTRY	
POPULATION	Scandinavian	Gaelic
Iceland	.375 (.215–.570)	.625 (.430–.785)
Orkney	.355 (.130–.645)	.645 (.355–.870)
Western Isles	.115 (.025–.355)	.885 (.645–.975)
Skye	.125 (.025–.405)	.875 (.595–.975)
NW Scottish coast	.135 (.035–.355)	.865 (.645–.965)

tended to be biased towards including genealogical links to high-ranking Norse ancestors is reflected in the following passage from the medieval text's introduction: "People often say that writing about the Settlements is irrelevant learning, but we think we can better meet the criticism of foreigners when they accuse us of being descended from slaves or scoundrels, if we know for certain the truth about our ancestry" (*The Book of Settlements* 1972, p. 6). According to the mitochondrial data, the truth seems to be that a sizeable portion of Icelandic lines of descent are traced back 1,100 years to females whose ancestry was firmly anchored in the British Isles.

In general, our findings indicate a good agreement between mtDNA variation and geography in Europe. In the case of the North Atlantic islands,   distances demonstrate that the most closely related mtDNA pools are geographic neighbors from the British Isles and Scandinavia. Genetic distances based on lineage-cluster frequencies demonstrate an even more marked and highly significant geographic structure of European mtDNA variation. These findings contrast with those of Simoni et al. (2000*a*), who detected only very limited geographic patterns in European mtDNA variation using spatial autocorrelation statistics (but see a different interpretation by Torroni et al. 2000). Although the methods are not strictly comparable, the very different conclusions reached by Simoni et al. (2000*a*) suggests the need for further examination of this problem. One potentially important difference between the two studies is that the average sample size in Simoni et al.'s (2000*a*) study was 73 (minimum 15 and maximum 249), whereas, in our analysis of genetic and geographic distances, the average sample size was 352 (minimum 128 and maximum 891). As shown in table 1, sampling saturation varies considerably among populations. Even for populations in which the sample size is

>   >   >>

landers and their African and Spanish ancestors inferred from mitochondrial DNA sequences. Ann Hum Genet 60: 321–330

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Paabo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837–840

Quintana-Murci L, Semino O, Bandelt HJ, Passarino G, McElreavey K, Santachiara-Benerecetti AS (1999) Genetic evidence of an early exit of *Homo sapiens sapiens* from Africa through eastern Africa. Nat Genet 23:437–441

Rafnson S (1999) The Atlantic islands. In: Sawyer P (ed) The Oxford illustrated history of the Vikings. Oxford University Press, Oxford

Richards MB, Macaulay VA, Bandelt HJ, Sykes BC (1998) Phylogeography of mitochondrial DNA in western Europe. Ann Hum Genet 62: 241–260

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Golge M, Dimitrov D, Hill E, Bradley D, Romano V, Cali F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, et al (2000) Tracing European founder lineages in the near eastern mtDNA