Am. J. Hum. Genet. 67:697–717, 2000

Estimating Scandinavian and Gaelic Ancestry in the Male Settlers of Iceland

Agnar Helgason,[1] Sigrún Sigurðardóttir,[3] Jayne Nicholson,[2] Bryan Sykes,[2] Emmeline W. Hill,[4] Daniel G. Bradley,[4] Vidar Bosnes,[5] Jeffery R. Gulcher,[3] Ryk Ward,[1] and Kári Stefánsson[3]

Institutes of [1]Biological Anthropology and [2]Molecular Medicine, University of Oxford, Oxford; [3]deCODE Genetics Inc., Reykjavik; [4]Department of Genetics, Trinity College, Dublin; and [5]Blodbanken, Ullevål Sykehus, Oslo

We present findings based on a study of Y-chromosome diallelic and microsatellite variation in 181 Icelanders, 233 Scandinavians, and 283 Gaels from Ireland and Scotland. All but one of the Icelandic Y chromosomes belong to haplogroup 1 (41.4%), haplogroup 2 (34.2%), or haplogroup 3 (23.8%). We present phylogenetic networks of Icelandic Y-chromosome variation, using haplotypes constructed from seven diallelic markers and eight microsatellite markers, and we propose two new clades. We also report, for the first time, the phylogenetic context of the microsatellite marker DYS385 in Europe. A comparison of haplotypes based on six diallelic loci and five microsatellite loci indicates that some Icelandic haplogroup-1 chromosomes are likely to have a Gaelic origin, whereas for most Icelandic haplogroup-2 and -3 chromosomes, a Scandinavian origin is probable. The data suggest that 20%–25% of Icelandic founding males had Gaelic ancestry, with the remainder having Norse ancestry. The closer relationship with the Scandinavian Y-chromosome pool is supported by the results of analyses of genetic distances and lineage sharing. These findings contrast with results based on mtDNA data, which indicate closer matrilineal links with populations of the British Isles. This supports the model, put forward by some historians, that the majority of females in the Icelandic founding population had Gaelic ancestry, whereas the majority of males had Scandinavian ancestry.

Introduction

Iceland was one of the last European landmasses to be colonized by humans. Historical sources (*The Book of Settlements* [1972]) record that the settlement of Iceland was orchestrated by Norse Vikings in 870–930 A.D., but they also indicate that individuals from the British Isles were among the founders. For more than half a century before Iceland was discovered and colonized, the Norse people had maintained settlements and wielded political power in the Shetlands, Orkneys, Hebrides, Isle of Man, and coastal regions of Ireland, Scotland, and northern England (Jones 1984; Davies 1999). Most historical and archaeological evidence suggests that, after initial periods of conflict, the Norse invaders, most of whom were men, settled and intermarried with existing populations (Sawyer 1997; Davies 1999). Later, when the colonization of Iceland commenced, many of these family groups subsequently left coastal settlements on the British Isles for a new life in the uncharted territory of Iceland. This would have resulted in a disproportionate number of females from the British Isles in the Icelandic founder population. Of the 48 women whose origin is referred to in the Icelandic *Book of Settlements* (1972), 16.7% have genealogical ties to the British Isles (Steffensen 1975). This contrasts with the finding that, of the 220 men for whom genealogical information is recorded in the same source, only 4.8% have British ancestry. It is known that the founders mentioned in *The Book of Settlements* (1972) represent only a small proportion of the total colonizing population, because it is thought that 8,000–20,000 individuals settled in Iceland in 870–930 A.D. (Steffensen 1975). A number of historians have argued that a disproportionate number of the individuals not mentioned in *The Book of Settlements* (1972) originated from the British Isles (Steffensen 1975; Sigurðsson 1988). Thus, for example, numerous slaves were captured by the Vikings in their raids on the coastlines of the British Isles, and many of the slaves were taken to Iceland. The majority of these slaves seem likely to have been female (Clover 1988; Karras 1988; Sayers 1994). In general, the expectation that males and females in the Icelandic settlement population may have had different levels of Scandinavian and British Isles ancestry underlines the importance of using sex-specific genetic markers to investigate the origin of Icelanders.

In a number of previous studies, investigators have

Received May 8, 2000; accepted for publication July 14, 2000; electronically published August 7, 2000.

Address for correspondence and reprints: Agnar Helgason, Institute of Biological Anthropology, University of Oxford, 58 Banbury Road, Oxford OX2 6QS, United Kingdom. E-mail: agnar.helgason@wolfson.ox.ac.uk

© 2000 by The American Society of Human Genetics. All rights reserved.
0002-9297/2000/6703-0017$02.00

attempted to assess the question of the relative proportions of Scandinavian and Gaelic individuals in the initial colonizing population, using a variety of historical, archaeological, linguistic, anthropometric, and genetic data. In initial studies of classical genetic markers (Bjarnason et al. 1973; Thompson 1973), a very substantial Gaelic component (>50%–98%) was found in the Icelandic settlement population. However, later studies, in which additional classical genetic markers were used, provided estimates ranging from 14% (Wijsman 1984) to 50% (Williams 1993). A recent study of mtDNA variation in Icelanders (Helgason et al. 2000) did not find a closer relationship to Scandinavian mitochondrial lineages, compared with those from the British Isles (the results of multidimensional scaling of analysis of molecular variance distances, based on sequences from hypervariable segment 1, indicated a closer affinity to the populations from the British Isles).

In the present study, we evaluated the diversity of Y-chromosome haplotypes in samples from Iceland, Norway, Sweden, Denmark, Ireland, and Scotland, using a combination of slowly evolving diallelic loci and rapidly evolving microsatellite loci. This allowed for an assessment of the relative diversity and phylogenetic context of the Icelandic Y-chromosome pool as well as an estimation of the proportions of Gaelic and Scandinavian male ancestry in the Icelandic gene pool.

Material and Methods

Population Samples, Loci, and Conditions

DNA samples were extracted from the blood of 181 unrelated Icelandic males who gave appropriate informed consent. Seven diallelic loci—92R7, M9, SRY-1532, YAP, TAT, LLY22g, and SRY-2627—were typed. 92R7 was amplified using the primers 5′-TGCATGAACACAAAAGACGTA-3′ and 5′-GCATTGTTAAATATGACCAGC-3′ and was digested with *Hin*dIII (Hurles et al. 1999). The primers and conditions used to obtain an amplicon flanking M9 were obtained from Underhill et al. (1997), and the site was typed by digestion with *Hin*fI. SRY-1532 was typed using the primers 5′-TCCTTAGCAACCATTAATCTGG-3′ and 5′-AAATAGCAAAAAATGACACAAGGC-3′ and was digested with *Dra*III (Santos et al. 1999). The presence or absence of the YAP *Alu* insertion was assayed using primers and conditions described elsewhere (Hammer and Horai 1995). The TAT marker was assayed using PCR with primers R5D and R5I and digestion with the restriction enzyme *Nla*III (Zerjal et al. 1997). LLY22g was typed using primers and conditions kindly supplied by C. Tyler-Smith (personal communication). SRY-2627 was typed using primers 5′-CGCGGCTTTGAATTTCAAGCTCTG-3′ and 5′-CCAGGGCCCCGAGGGACTCTT-3′ and was digested with *Ban*I (Hurles et al. 1999).

Nine Y-chromosome microsatellite loci were typed using fluorescently labeled primers, including six tetranucleotide repeat markers (DYS19, DYS389I, DYS389II, DYS390, DYS391, and DYS393), two trinucleotide repeat loci (DYS388 and DYS392), and one tetranucleotide repeat marker (DYS385, which provides two Y chromosome–specific alleles). The primers used were those described by Kayser et al. (1997). PCRs were set up, run, and pooled on Applied Biosystems 877 integrated catalyst thermocyclers (PE Biosystems). Multiplex 5-μl reactions, each of which used 10 ng template genomic DNA, 0.25 U Amplitaq Gold (PE Biosystems), 2 pmol each primer, 0.2 mM deoxynucleotides, and 2.5 mM $MgCl_2$, were set up; the buffer was supplied by PE Biosystems. The following PCR conditions were used: activation of Amplitaq Gold at 95°C for 10 min, 12 cycles of denaturation at 95°C for 15 s, annealing at 55°C for 30 s, and elongation at 72°C for 1 min, followed by 28 cycles at 89°C for 15 s, annealing at 55°C for 30 s, and elongation at 72°C for 1 min. After supplementation with internal size standards, the products were electrophoresed on an Applied Biosystems model 377 sequencer and were detected using TrueAllele (Cybergenetics). The genotypes were defined and edited with the use of Decode-GT 1.0 software (Pálsson et al. 1999).

DNA from 61 Scottish males was extracted from blood samples collected at Blood Transfusion Service donor sessions throughout Scotland. Similarly, DNA from 110 Norwegian males was extracted from blood samples collected at donor sessions at the Ullevål and Red Cross hospitals in Oslo. In both instances, individuals gave informed consent, and information about paternal ancestry indicated that the samples were geographically representative for all geographic areas of Scotland and Norway, respectively. Six diallelic loci (92R7, M9, SRY-1532, YAP, TAT, and SRY-2627) were typed, together with five microsatellite loci (DYS19 and DYS390–DYS393). The primers and conditions used for the microsatellite loci and the diallelic loci M9, SRY-1532, YAP, and SRY-2627 are described elsewhere (Hurles et al. 1998), and those used for 92R7, TAT, and LLY22g were the same as those described for Icelandic samples.

In addition, 222 Irish Y-chromosome haplotypes (221 chromosomes described by Hill et al. [2000] and 1 chromosome described by Hurles et al. [1999]) were included for comparative analysis, as were 12 Danish Y-chromosome haplotypes and 109 Swedish Y-chromosome haplotypes (T. Zerjal, L. Beckman, and C. Tyler-Smith, personal communication). Each of these samples was typed for six diallelic loci (92R7, M9, SRY-1532, YAP, TAT, and SRY-2627) and five microsatellite loci (DYS19 and DYS390–DYS393). In much of the following anal-

ysis, Norwegians, Swedes, and Danes are collectively referred to as Scandinavians, whereas Irish and Scottish populations are referred to as Gaels. Data are presented in tables1 and2.

Summary Statistics

All the loci map to the nonrecombining part of the Y chromosome and can be used to construct combined haplotypes. Except where stated otherwise, the common set of loci used for all statistical comparisons of Icelandic, Scandinavian, and Gaelic Y-chromosome haplotypes included six diallelic loci (92R7, M9, SRY-1532, YAP, TAT, and SRY-2627) and five microsatellite loci (DYS19 and DYS390–DYS393). Gene diversity was estimated by: $(n/n-1)(1-\Sigma_{i=1}^{k}p_i^2)$, where n is the sample size, k is the number of distinct haplotypes, and p_i is the frequency of each haplotype. This index represents the probability that two randomly chosen haplotypes from a sample are different.

Because the distribution of Y-chromosome haplotypes defined by a combination of diallelic and microsatellite loci will approximate that of an infinite allele model, we used Ewens' sampling formula (Ewens 1972) to estimate the population mutation parameter θ_k as $E(k) = \theta_k \Sigma_{i=0}^{n-1}(1/\theta_k + i)$, where k is the number of distinct haplotypes observed in a sample size of n and where θ_k denotes $2N_{me}\mu$, for which N_{me} denotes the male effective population size and μ denotes the rate at which new haplotypes are generated. Because μ can be considered to be constant across populations, variation in θ_k should reflect differences in N_{me}.

The frequency spectrum of Y-chromosome haplotypes in populations with moderate θ_k is characterized by a large number of rare haplotypes. In such cases, sample sizes must be large to ensure adequate haplotype detection. Given that inadequate levels of sampling saturation can confound estimates of genetic diversity, genetic distance, and admixture (Pfeiffer et al. 1999; Helgason et al. 2000), we also used Ewens' sampling formula (Ewens 1972) to estimate the adequacy of each sample. For a given θ_k, we arbitrarily consider an adequate sample to be one for which an increase of 10 individuals is expected to reveal less than one new haplotype (Helgason et al., in press). Division of the actual sample size by this value provides an index of the sample saturation ratio.

The diversity of haplotypes within populations and haplogroups was estimated as the average squared mutational difference (ASD) between all pairs of haplotypes: $\text{ASD} = \Sigma_{i=1}^{k}\Sigma_{j<i}p_ip_jd_{ij}^2$. Under the assumption of a stepwise model of microsatellite mutation, d_{ij} is the number of differences between haplotypes i and j in a population, k is the number of distinct haplotypes, and p_i and p_j are the frequencies of haplotypes i and j, respectively. For ASD calculations involving haplotypes constructed from both diallelic and microsatellite loci, diallelic mutations were weighted 1,000-fold higher than microsatellite mutations. This disparity is intended to reflect the very different rates at which these two types of loci mutate.

ASD can provide a useful summary of the mutational divergence between all pairs of haplotypes within a population and between all pairs of haplotypes from two populations. However, in the context of our study, a strict interpretation of ASD is inappropriate—that is, the diversity and divergence of the Icelanders, Scandinavians, and Gaels, as measured by ASD, have not arisen exclusively through mutation events over a long period of time within these populations. Instead, their Y-chromosome pools have been shaped primarily by gene flow and genetic drift, interspersed with some mutation events (see Calafell et al. [2000] for similar arguments regarding autosomal microsatellite variation in human populations). This is self-evident in the case of the Icelanders, a population established by a recent large-scale migration, but it is also true for Scandinavians and Gaels, as demonstrated by the considerable degree of Y-chromosome haplotype sharing among these populations (see the Results section). The actual mutational divergence between these populations is most likely to have occurred at the rapidly evolving microsatellite loci. To more effectively summarize the portion of mutational diversity, both within and between populations, that is relevant to the population histories of the Icelanders, Scandinavians, and Gaels, we calculated ASD between all pairs of haplotypes within and between populations separately for each haplogroup. We then summarized these haplogroup-specific values for populations through the use of averages weighted by the total number of haplotypes in each haplogroup.

Two other measures of genetic distance between populations were estimated. The first measure, haplotype divergence, is based on the genetic chord distance, as proposed by Cavalli-Sforza and Edwards (1967). The divergence for a single locus is given by $f_\theta = 4(1 - \Sigma_{i=1}^{k}\sqrt{p_{iA}p_{iB}})/(k-1)$, where p_{iA} and p_{iB} are the frequencies of the ith haplotype in populations A and B and where k is the total number of haplotypes. This divergence measure assumes that population differentiation is primarily due to genetic drift. The second measure (ρ) is defined as the average number of mutations between the haplotypes of one population and the closest founder haplotypes observed (or, in some applications of the method, reconstructed) in another population (see Forster et al. 1996). This genetic distance measures the minimum mutational divergence between the haplotypes of two gene pools.

Table 1

Icelandic Y-Chromosome Haplotypes

Haplo-group[a]	DYS19	DYS390	DYS391	DYS392	DYS393	DYS388	DYS389I	DYS389b	DYS385	No. of Individuals
1	13	23	10	12	13	12	9	16	13*17	1
	13	23	10	12	13	12	9	17	13*17	5
	13	23	10	12	13	12	9	17	13*18	1
	13	23	10	12	13	12	10	16	13*17	1
	13	23	11	12	13	12	9	17	13*17	3
	13	23	11	12	13	12	9	18	13*17	1
	13	24	10	13	13	12	10	16	11*14	1
	13	24	10	13	13	12	10	16	12*14	1
	14	22	10	13	13	12	11	16	11*14	1
	14	23	9	13	13	12	10	16	11*14	1
	14	23	9	13	13	12	10	17	11*14	1
	14	23	10	13	13	12	10	15	15*16	1
	14	23	10	13	13	12	10	17	11*14	1
	14	23	11	13	13	12	10	16	11*11	1
	14	23	11	13	13	12	10	16	11*14	2
	14	23	11	13	13	12	11	16	11*14	3
	14	23	12	13	13	12	10	16	11*16	1
	14	24	10	13	13	12	9	16	11*15	1
	14	24	10	13	13	12	10	15	11*14	1
	14	24	10	13	13	12	10	16	11*14	3
	14	24	10	13	13	14	10	15	11*14	1
	14	24	11	13	12	12	10	16	11*13	1
	14	24	11	13	13	12	10	15	11*13	1
	14	24	11	13	13	12	10	15	11*14	2
	14	24	11	13	13	12	10	16	11*14	8
	14	24	11	13	13	12	10	16	11*15	1
	14	24	11	13	13	12	10	16	12*14	1
	14	24	11	13	13	12	10	16	11*13	5
	14	24	11	13	13	12	10	17	11*14	1
	14	24	11	13	13	12	11	16	11*14	1
	14	24	11	13	13	12	11	16	11*13	2
	14	24	11	13	13	12	11	16	12*14	1
	14	24	11	14	13	12	10	16	11*13	1
	14	24	12	13	13	12	11	16	11*14	1
	14	24	12	13	14	12	11	16	11*14	1
	14	24	13	13	13	12	11	16	11*14	1
	14	25	10	13	13	12	10	17	11*14	1
	14	25	11	13	13	11	10	16	11*14	1
	14	25	11	13	13	12	10	16	11*14	5
	14	25	11	13	13	12	10	16	12*15	1
	14	25	11	14	13	12	10	14	11*13	1
	14	25	11	14	13	12	10	16	11*13	1
	14	25	11	14	13	12	11	15	11*13	1
	14	26	11	13	13	11	10	16	11*14	2
	14	27	11	13	13	12	10	16	11*14	1
	15	24	11	13	14	12	10	16	11*13	1
16	14	25	10	14	14	12	11	15	11*13	1
2	14	22	10	11	13	14	9	16	13*14	1
	14	22	10	11	13	14	9	16	14*15	1
	14	22	10	11	13	14	9	16	14*14	1
	14	22	10	11	13	14	10	17	14*14	1
	14	23	10	11	12	14	9	16	15*15	1
	14	23	10	11	13	12	10	15	14*15	1
	14	23	10	11	13	14	9	15	14*15	1
	14	23	10	11	13	14	9	16	13*14	3
	14	23	10	11	13	14	9	16	13*15	5
	14	23	10	11	13	14	9	16	14*14	3
	14	23	10	11	13	14	9	16	14*15	21

(*continued*)

Table 1 (Continued)

Haplo-group[a]	DYS19	DYS390	DYS391	DYS392	DYS393	DYS388	DYS389I	DYS389b	DYS385	No. of Individuals
	14	23	10	11	13	14	9	16	15*15	1
	14	23	10	11	13	14	9	17	14*14	1
	14	23	10	11	13	14	10	16	13*14	1
	14	23	10	11	13	14	10	16	14*15	1
	14	23	10	11	14	14	9	16	14*14	1
	14	23	10	12	13	14	10	16	14*15	1
	14	24	10	11	13	12	9	16	13*14	1
	14	24	10	11	13	15	10	16	13*15	1
	15	22	10	11	13	14	9	16	13*14	1
	15	22	10	11	14	13	9	18	14*14	1
	15	23	10	11	13	14	9	16	14*14	1
	15	23	10	11	14	14	9	16	14*14	1
	15	23	10	12	15	13	10	17	15*16	3
	15	23	11	11	13	14	9	16	14*15	1
	16	22	10	11	13	14	9	16	13*14	1
	16	23	10	11	13	13	11	16	12*12	3
	16	23	10	11	14	13	11	18	14*15	1
	16	26	11	12	14	13	11	16	15*15	1
	17	22	10	11	13	14	9	16	13*14	1
3	14	24	11	11	13	12	10	17	11*13	1
	15	22	11	11	13	12	10	18	11*14	1
	15	24	11	11	13	12	10	17	11*14	2
	15	25	10	11	13	12	9	16	11*14	2
	15	25	10	11	13	12	10	16	11*14	2
	15	25	10	11	13	12	11	17	11*14	1
	15	25	10	11	13	12	11	18	11*14	1
	15	25	11	11	13	12	9	16	11*14	4
	15	25	11	11	13	12	10	17	11*14	4
	15	25	11	11	13	12	10	17	11*15	2
	15	25	11	11	13	12	10	18	11*14	1
	15	25	11	11	13	12	11	17	11*14	6
	15	25	11	11	13	12	11	18	11*14	2
	15	25	11	11	13	12	11	18	11*13	1
	15	25	11	11	13	12	12	17	11*14	1
	15	27	10	11	13	12	10	17	11*14	1
	16	24	11	11	13	12	10	17	11*14	1
	16	25	10	11	13	12	10	17	11*14	1
	16	25	10	11	13	12	11	18	11*14	1
	16	25	11	9	13	12	10	16	11*14	1
	16	25	11	11	13	12	10	16	11*14	4
	16	25	11	11	13	12	11	17	11*14	2
	16	25	12	11	13	12	10	16	11*14	1

[a] The diallelic loci character state for haplogroup 1 is 1110000; for haplogroup 2, 0010000; for haplogroup 3, 1100000; and for haplogroup 16, 0111100. The order of the diallelic loci is 92R7, M9, SRY-1532, LLY22g, TAT, YAP, and SRY-2627.

Phylogenetic Reconstruction

Phylogenetic relationships between Icelandic microsatellite haplotypes within haplogroups, as defined by the diallelic loci, were reconstructed in the form of median-joining networks, with use of the program Network 2.0 (Bandelt et al. 1999; software available at the Life Sciences and Engineering Technology Solutions Web site). For this part of the analysis, the larger set of microsatellite loci (including DYS388, DYS389I, DYS389II, and DYS385) typed for the Icelanders was used. Because the repeat length of DYS389II includes DYS389I, the latter was subtracted from DYS389II to derive DYS389b. A weighting scheme was used to reflect differences in the mutation-rate estimates, as reported by Kayser et al. (2000), whereby DYS19, DYS388, DYS389I, DYS392, and DYS393 were given weights of 3, DYS389b and DYS391 were given weights of 2, and DYS390 was given a weight of 1. We use the estimated mutation rates, rather than the mean repeat number of alleles (as proposed by Forster et al. [2000]), because, with the exception of DYS390, the correlation between mutation rate and mean allele size is low (with DYS390, $r = 0.72$; without DYS390, $r = 0.07$). The mean allele sizes of microsatellite loci in the Icelanders are 14.48,

Table 2

Y-Chromosome Haplotypes in Norwegians, Danes, Scots, and Irish

Haplo-group	DYS19	DYS390	DYS391	DYS392	DYS393	No. of Individuals			
						Norwegians	Danes	Irish	Scots
1	13	23	10	13	14	1			
	13	23	10	14	13	1			
	13	23	11	13	13			1	
	13	23	11	14	13	1			
	13	24	10	14	13	1			
	14	21	11	13	13			1	
	14	22	10	13	13			1	
	14	22	11	13	13			2	
	14	23	10	13	13	2	1	9	2
	14	23	10	13	14			1	
	14	23	11	12	13	1			
	14	23	11	13	12			1	
	14	23	11	13	13	2		14	3
	14	23	11	13	14	1			
	14	23	11	14	13			1	
	14	23	11	15	13			1	
	14	24	9	14	13			1	
	14	24	10	13	13	3	1	26	13
	14	24	10	13	14			1	
	14	24	10	14	13			5	2
	14	24	11	12	13	1			
	14	24	11	12	14			1	
	14	24	11	13	12			2	1
	14	24	11	13	13	6	2	35	11
	14	24	11	13	14				2
	14	24	11	14	13	1		5	
	14	24	12	13	13	1		1	
	14	24	12	13	14			1	
	14	25	10	13	12				2
	14	25	10	13	13			2	1
	14	25	10	14	13			8	
	14	25	10	15	13			1	
	14	25	11	12	13			1	1
	14	25	11	13	13	2	1	14	2
	14	25	11	14	13	1		23	
	14	25	11	14	14			1	
	14	25	11	15	13			1	
	14	25	12	14	13			1	
	14	26	10	14	13			1	
	14	26	11	13	13	1			1
	14	26	11	14	13			4	
	15	23	10	13	13			3	
	15	23	11	13	13			1	
	15	23	11	13	15			1	
	15	24	10	13	13	1		2	1
	15	24	11	13	13	2		2	2
	15	24	11	13	14			1	
	15	24	12	13	13				1
	15	25	11	13	13				1
	15	25	11	14	13			3	
	16	24	11	13	13			1	
	17	23	11	12	13				1
16	14	23	10	14	14	2			
	14	23	11	14	14	1			
	15	23	11	14	14			1	

(*continued*)

Table 2 (Continued)

Haplo-group	DYS19	DYS390	DYS391	DYS392	DYS393	No. of Individuals			
						Norwegians	Danes	Irish	Scots
2	13	23	10	11	13	1			
	14	21	10	12	13	1			
	14	21	11	11	13			1	
	14	22	10	11	12			2	
	14	22	10	11	13	11		3	1
	14	22	10	11	14			2	
	14	23	10	11	12			1	
	14	23	10	11	13	21		5	5
	14	23	10	12	13		1		
	14	23	10	12	15			1	
	14	23	11	11	13		1		
	14	24	10	11	12	1			
	14	24	10	11	13	1			
	15	22	10	11	12	2			
	15	22	10	11	13	6		1	1
	15	22	10	11	14			1	
	15	22	10	12	13	1			
	15	23	9	11	12	1			
	15	23	10	11	12	1			
	15	23	10	11	13			3	
	15	23	10	12	14			2	
	15	23	10	12	15			4	
	15	23	11	11	13			1	
	15	23	11	12	15		1	1	
	15	24	10	11	13			1	
	15	24	10	12	14			1	
	15	24	10	12	15			1	
	15	24	11	12	13			1	
	16	22	10	11	13	8			
	16	23	9	11	12		1		
	16	23	10	12	14				1
	16	23	12	12	14	1			
	16	24	10	11	13			1	
	17	22	10	11	13	1			
26	12	23	10	14	13			1	
3	15	24	10	11	13	2			
	15	24	11	11	13	2			
	15	25	10	11	13	1			1
	15	25	11	11	13	6		1	1
	15	25	11	11	14	1			
	15	25	12	11	13	1			
	15	26	11	11	13	1			
	16	24	10	11	13	1			
	16	25	10	11	13	2			
	16	25	10	11	14	1			
	16	25	11	11	13	1			1
	16	25	12	11	13		1		
	17	24	10	11	13	1			
	17	25	10	11	13				1
	17	25	11	11	13		1		
4/21	13	23	11	11	13			1	
	13	24	10	11	13	2	1	1	2
	13	24	10	11	14			1	
	13	25	9	11	14			1	
	14	24	10	11	13	1		1	

Note.—The diallelic loci character states for each haplogroup are as follows: for haplogroup 1, 111000; for haplogroup 2, 001000; for haplogroup 3, 110000; for haplogroup 4/21, 001010; for haplogroup 16, 011100; and for haplogroup 26, 011000. The order of diallelic loci is 92R7, M9, SRY-1532, TAT, YAP, and SRY-2627.

for DYS19; 12.60, for DYS388; 9.85, for DYS389I; 16.34, for DYS389b; 23.91, for DYS390; 10.53, for DYS391; 11.64, for DYS392; and 13.06, for DYS392.

Median-joining networks were generated with ε values of 0, because this yields a network containing a limited number of plausible evolutionary pathways among haplotypes. Setting the $\varepsilon > 0$ increases the number of reconstructed median vectors, which increases the probability that the true phylogenetic tree is contained within the network (Bandelt et al. 1999) but which also leads to a vast increase in the number of false trees displayed in the network. We report on specific features of the networks obtained when $\varepsilon > 0$.

Admixture Estimates

Most admixture estimators are based on Bernstein's model (1931), in which allele frequencies in the admixed population are considered to be linear combinations of allele frequencies from the source populations and the effects of drift and mutation are presumed to be negligible. In this model, allele frequencies in the admixed population are given by $p_h = p_1 m + p_2(1 - m)$, where m is the admixture proportion of the first source population and where p_h, p_1, and p_2 are allele frequencies in the admixed population and the two source populations, respectively. We used the Admix1_0 software program (Bertorelle and Excoffier 1998) to obtain the standard least-squares estimate for this model. This estimate, m_R is based on fitting a regression line through the points defined by $p_1 - p_2$ and $p_h - p_2$ (Roberts and Hiorns 1965).

However, there are drawbacks to application of the standard model to Y-chromosome haplotype data. The multinomial sampling distribution, which is appropriate for a classical k-allele system, is an inadequate approximation for the sampling distribution of Y-chromosome haplotypes that exhibit a frequency spectrum of an infinite allele model with moderate θ_k. In addition to the large sampling variation associated with the infinite allele model, the effects of drift and mutation also result in the admixed sample containing a large number of "private" haplotypes. These cannot be incorporated into the standard model, because $p_1 = p_2 = 0$; we used two additional admixture estimators to overcome these deficiencies of the standard model.

To overcome the problems caused by the excessive variance of the infinite allele model, we devised a heuristic iterative approach to define the distribution of the admixture estimate that best fit the data. This estimator, designated as m_ρ, is designed to incorporate information from the private Icelandic haplotypes and is obtained as follows. Given a prior probability of admixture, η, the probability that a randomly chosen haplotype observed in the Icelandic 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 haplotype in the two source populations, as before. For a given value of η, 10,000 random samples are used to obtain the mean and 95% confidence intervals for the admixture estimate m_ρ, conditioned on the prior value of η. The fit to the data is evaluated by $\Sigma(Nm_\rho - N\eta)^2/N\eta$, where N is the size of the Icelandic sample. The best-fitting model is found by iteratively searching over the line $0 \leqslant m_\rho \leqslant 1$ in successively smaller intervals around the best-fitting value for m_ρ. The 95% confidence intervals (CIs) were estimated by simulation. The putative founder haplotype(s) for a private Icelandic haplotype is defined as the haplotype in the source population(s) that differs by the smallest number of mutations, under the assumption of a stepwise mutation for microsatellites and with use of the distance matrix ρ, as defined in the Summary Statistics section. If more than one haplotype in the source population(s) meets this criterion as a putative founder haplotype (as in the case of a tie), then their cumulative frequency is used to derive p_1 and p_2, to calculate the conditional probability of origin.

Like the m_R least-squares admixture estimator, the m_ρ strategy is based on haplotype frequencies, but it also incorporates the effect of mutation that may have occurred since the admixture event. To more explicitly model the potential effect of mutational divergence, we used the m_Y estimator proposed by Bertorelle and Excoffier (1998) and calculated with use of their Admix1_0 software program. The difficulty associated with application of this estimator to the Icelandic situation is that, because $\tau < 0.1$ and $t_A < 0.01$, the model will tend to be biased toward a 50:50 estimate of admixture (τ is the divergence time between the parental populations, and t_A is the time since the admixture event; both are measured in units of $1/(2u)$ generations, where u is the mutation rate) (Bertorelle and Excoffier 1998). Moreover, because it is based on a matrix of ASD distances between Y-chromosome haplotypes, m_Y will be biased by the underlying phylogeny of the heavily weighted diallelic loci (see the Summary Statistics section). As noted in the Results section, this estimator provided quite different results from the other two methods.

Results

Haplogroup Frequencies

The diallelic loci used in this study partition European Y chromosomes into eight clades, which are usually referred to as "haplogroups" (see fig. 1). A number of studies have demonstrated nonrandom geographic patterns of Y-chromosome haplogroup distribution in Europe (Zerjal et al. 1997; Hurles et al. 1999; Lucotte and Loirat 1999; Hill et al. 2000). Figure 2 shows a map of

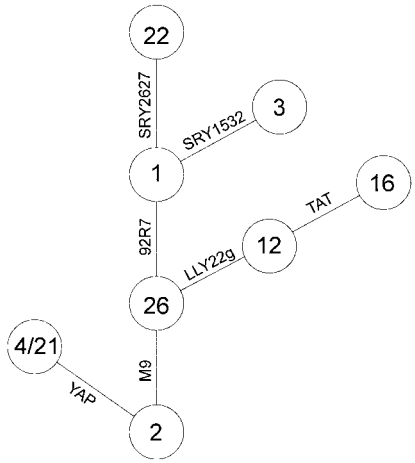

Figure 1 Haplogroups are represented by circles and are identified by numbers, on the basis of the nomenclature of Tyler-Smith (1999) and Jobling et al. (1997). Lines denote mutation events, which are labeled with the associated diallelic marker.

haplogroup frequencies (as defined by the diallelic loci shown in fig. 1) for the Icelanders, Norwegians, Swedes, Scots, Irish, and a number of other European populations. The haplogroup frequencies are also listed in table 3. The most striking feature of these haplogroup distributions is the difference between the populations of the British Isles and those of Scandinavia. As reported elsewhere (Hill et al. 2000), haplogroup 1 is predominant in Ireland, and this is also the case in Scotland and England. In contrast, Scandinavian populations have a considerably lower frequency of haplogroup 1 but higher frequencies of both haplogroups 2 and 3. These differences between Scandinavia and the British Isles indicate that Y-chromosome haplotypes should be informative for detecting the result of male admixture in the Icelandic population. Thus, the level of Gaelic admixture in the Icelanders is likely to be reflected by higher frequencies of haplogroup 1 and lower frequencies of haplogroups 2 and 3, compared with those observed in Scandinavia. This is more or less what we find in the Icelanders. The frequencies of haplogroups 1 and 2 (41.8% and 34.6%, respectively) in the Icelanders are intermediate between those of the same haplogroups in populations from Scandinavia and the British Isles. Haplogroup 3, however, is more frequent in Icelanders than in Scandinavians or Gaels.

Phylogenetic Networks and Haplotype Sharing

Figure 2 shows that frequencies of haplogroups 1 and 2 in Icelanders are intermediate between the values in Gaels and Scandinavians, and it suggests an admixed Icelandic population with Gaelic haplogroup-1 chromosomes and Scandinavian haplogroup-2 and -3 chromosomes. However, note that 14% of Gaelic Y chromosomes belong to haplogroup 2 and that 24% of Scandinavian chromosomes belong to haplogroup 1, indicating the possibility of a Gaelic contribution to Icelandic haplogroup-2 chromosomes and a Scandinavian contribution to Icelandic haplogroup-1 chromosomes. Moreover, the high frequency of haplogroup-3 chromosomes in the Icelanders raises the issue of whether their haplogroup frequencies could have been altered by the effects of genetic drift (see Guldberg et al. 1997; Helgason et al. 2000).

When the rapidly mutating microsatellite loci are combined with the more stable diallelic loci, the Y-chromosome haplogroups become partitioned into a great variety of haplotypes. This abundant genetic variation allows for a more precise identification of ancestral links between the Y chromosomes of the Icelanders and their putative source populations. At the same time, the deep phylogenetic resolution of the diallelic loci reduces the problem of recurrent mutations at microsatellite loci, yielding haplotypes that are identical by state but not by descent. Figures 3, 4, and 5 show median-joining networks describing the phylogenetic relationships between Icelandic haplotypes reconstructed on the basis of all the microsatellite loci (except DYS385) within haplogroups 1, 2, and 3, respectively. Because the primers for DYS385 yield two alleles of similar size, they cannot be used for phylogenetic reconstruction in the same way that other single-allele loci are used. However, the allele sizes for the DYS385 microsatellite have been included post hoc and are shown below each haplotype in figures 3, 4, and 5. Because DYS385 alleles are highly variable, they can be used to independently assess the integrity of reconstructed haplotype clusters.

Not surprisingly, the broad pattern revealed by the haplogroup frequencies also appears to hold for the haplotypes constructed from both diallelic loci and microsatellite loci. However, the association between haplogroups and source populations is not uniform. For haplogroups 1 and 2, the majority of Icelandic haplotypes are found in both Scandinavians and Gaels. Figure 3 shows that haplogroup-1 chromosomes are more frequent in the Gaels, as would be expected from the haplogroup frequencies in figure 2. Four haplotypes are shared exclusively by Icelanders and Gaels—the only haplogroup in which this occurs. However, two haplotypes, both of which belong to a long branch in haplogroup 1 (fig. 3, *branch A*), are also shared exclusively by Iceland and Scandinavia. The length of this branch suggests that it may represent a separate haplogroup for which a diagnostic diallelic marker has not yet been identified. This interpretation is supported by the fact that, in spite of their prevalence in haplogroup 1, no Gaelic haplotypes are found on this branch. Also, the five haplotypes on branch A all have DYS385 alleles

13*17 or 13*18, which are distinct from the DYS385 alleles (typically, 11*14) in the other haplotypes in figure 3. Furthermore, the two haplotypes shared by Icelanders and Scandinavians at the end of branch A exhibit an increase in allele size (from 12 to 13 repeats) for the microsatellite DXYS156Y (data not shown). The integrity of branch A was retained in network reconstructions with ε varying from 0 to 4, whereas other branches were increasingly transformed into reticulations for this range of ε.

Figure 4 shows a tightly linked cluster of haplogroup-2 chromosomes with a preponderance of Scandinavian chromosomes and five haplotypes exhibiting substantial evolutionary divergence from the rest. The divergence of these haplotypes is unlikely to be due to microsatellite-typing errors, because three of the five haplotypes are also observed in Scandinavians and/or Gaels. Moreover, the most divergent haplotype has been confirmed by independent typing of the same loci in that individual's father, and the microsatellite profile of the other Icelandic private haplotype is shared (with a difference of one repeat the DYS389b locus) by a Norwegian (see the Y-STR Haplotype Reference Database). It is most likely that these haplotypes belong to one or more separate haplogroups that are not defined by the diallelic loci used

Table 3

Haplogroup Frequencies in European Populations

		Frequency of Haplogroup (%)					
Population	*n*	1	2	3	4/21	16	26
Icelanders	181	41.44	34.25	23.76	.00	.55	.00
Norwegians	112	25.89	50.89	17.86	2.68	2.68	.00
Swedes	110	20.00	53.64	17.27	.91	7.27	.91
Danes	12	41.67	33.33	16.67	8.33	.00	.00
Irish	222	81.53	14.86	.45	2.25	.45	.45
Scots	61	77.05	13.11	6.56	3.28	.00	.00
British	32	68.75	18.75	9.38	3.13	.00	.00
Germans	32	46.88	34.38	9.38	6.25	.00	3.13
Greeks	42	14.29	45.24	4.76	33.33	2.38	.00
Italians	332	35.84	45.48	2.71	11.14	.00	4.82
Russians	30	26.67	16.67	43.33	3.33	3.33	6.67

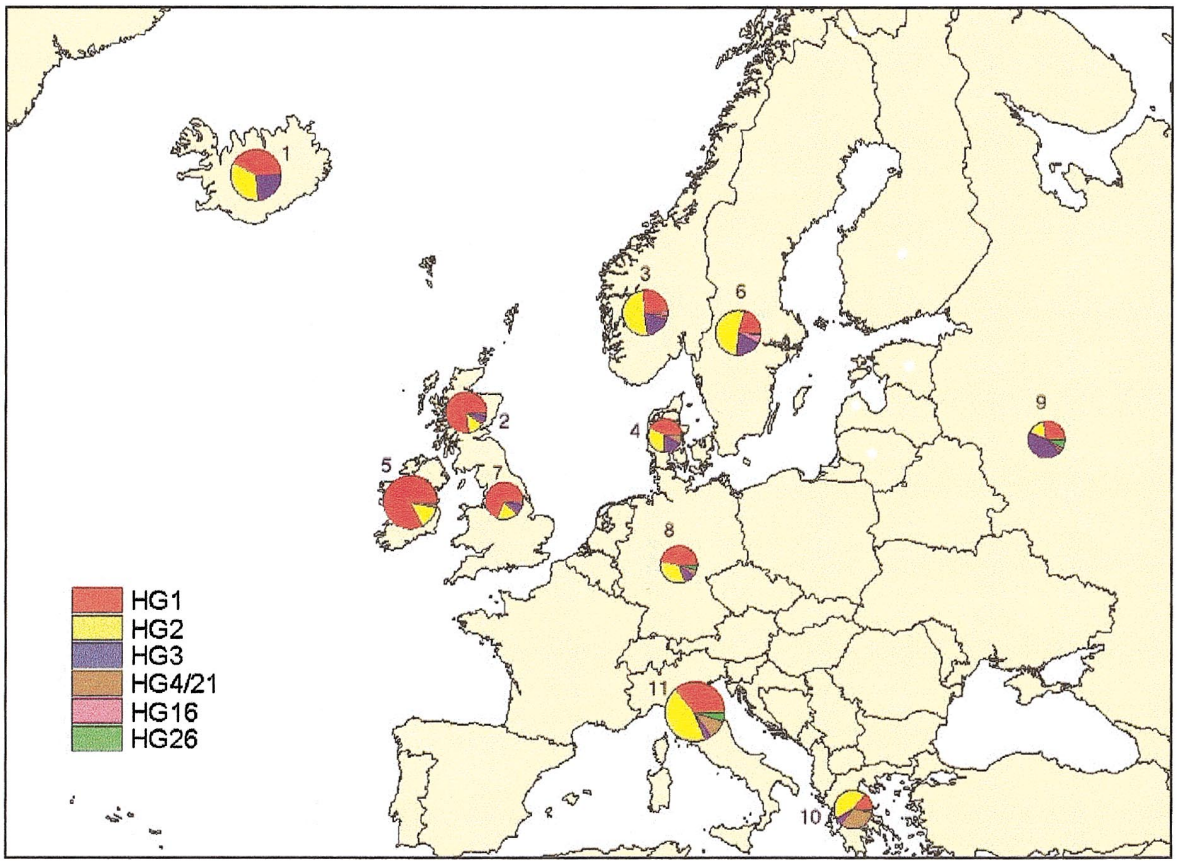

Figure 2 The area of each pie denotes sample size, and the pie slices denote the proportion of chromosomes belonging to each haplogroup. Haplogroups were assigned using the diallelic markers 92R7, M9, SRY-1532, YAP, and TAT. Sources of data are as follows: for pies 1–4, the present study; for pie 5, Hill et al. (2000); for pie 6, T. Zerjal, L. Beckman, and C. Tyler-Smith (personal communication); for pies 7–10, Karafet et al. (1999); and for pie 11, Previdere et al. (1998).

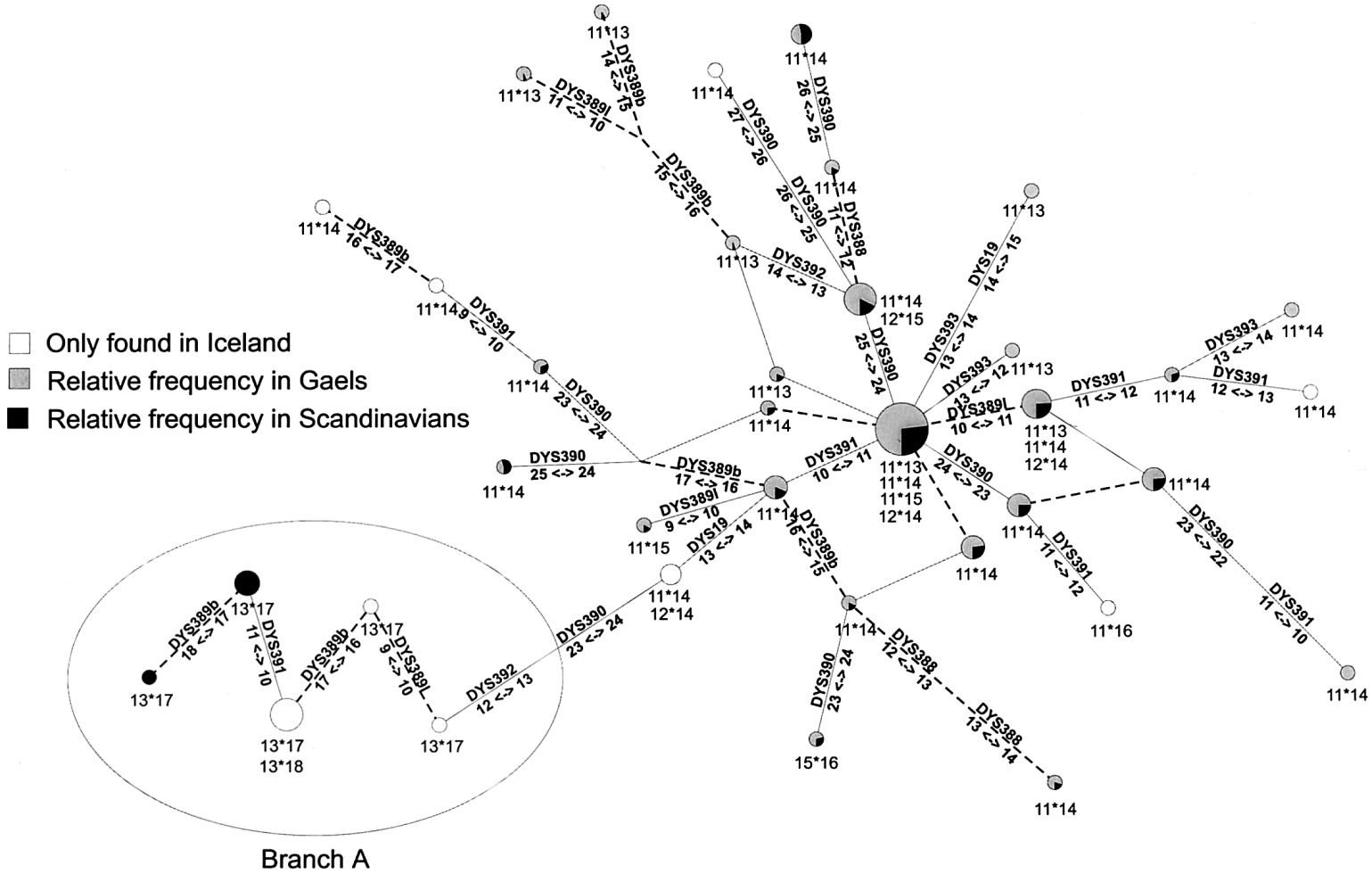

Figure 3 Median-joining network showing the phylogenetic relationships of Icelandic Y-chromosome haplotypes within haplogroup 1. Haplogroup membership was determined by the seven diallelic markers listed in the Material and Methods section. Haplotypes are defined by character states at eight microsatellite loci: DYS19, DYS388, DYS389I, DYS389b, and DYS390–DYS393. The allele sizes of the two DYS385 loci are shown below each haplotype. Circles denote haplotypes, and their sizes denote the frequencies of the haplotypes in the Icelanders. Lines denote mutational changes in the repeat numbers of microsatellite loci, with both the locus and the change in repeat number indicated. Reticulations denote the presence of recurrent mutations that could not be resolved with the available markers. Circles are shaded with respect to haplotype sharing with Scandinavians and Gaels, on the basis of a comparison of a restricted set of five microsatellites: DYS19 and DYS390–DYS393. Broken lines denote mutations at loci that are not included in the restricted set of microsatellites and therefore do not discriminate between haplotypes of Icelanders and those of Scandinavians and Gaels. Gray-shaded and blackened areas denote the relative frequencies of a haplotype in Gaels and Scandinavians, respectively. Unshaded and unblackened haplotypes are found only in Iceland.

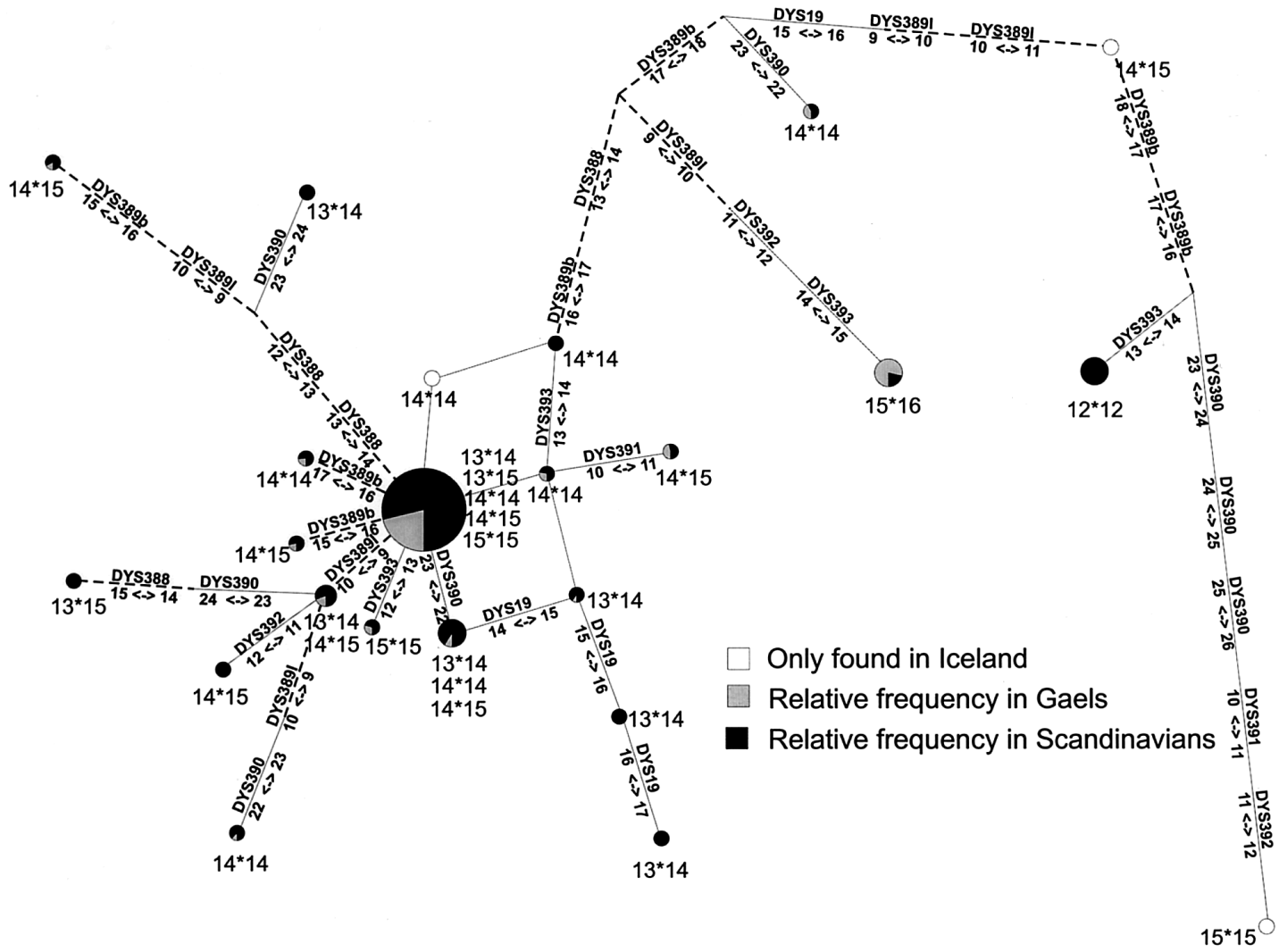

Figure 4 Median-joining network showing the phylogenetic relationships of Icelandic Y-chromosome haplotypes within haplogroup 2. For definitions, see the legend for figure 3.

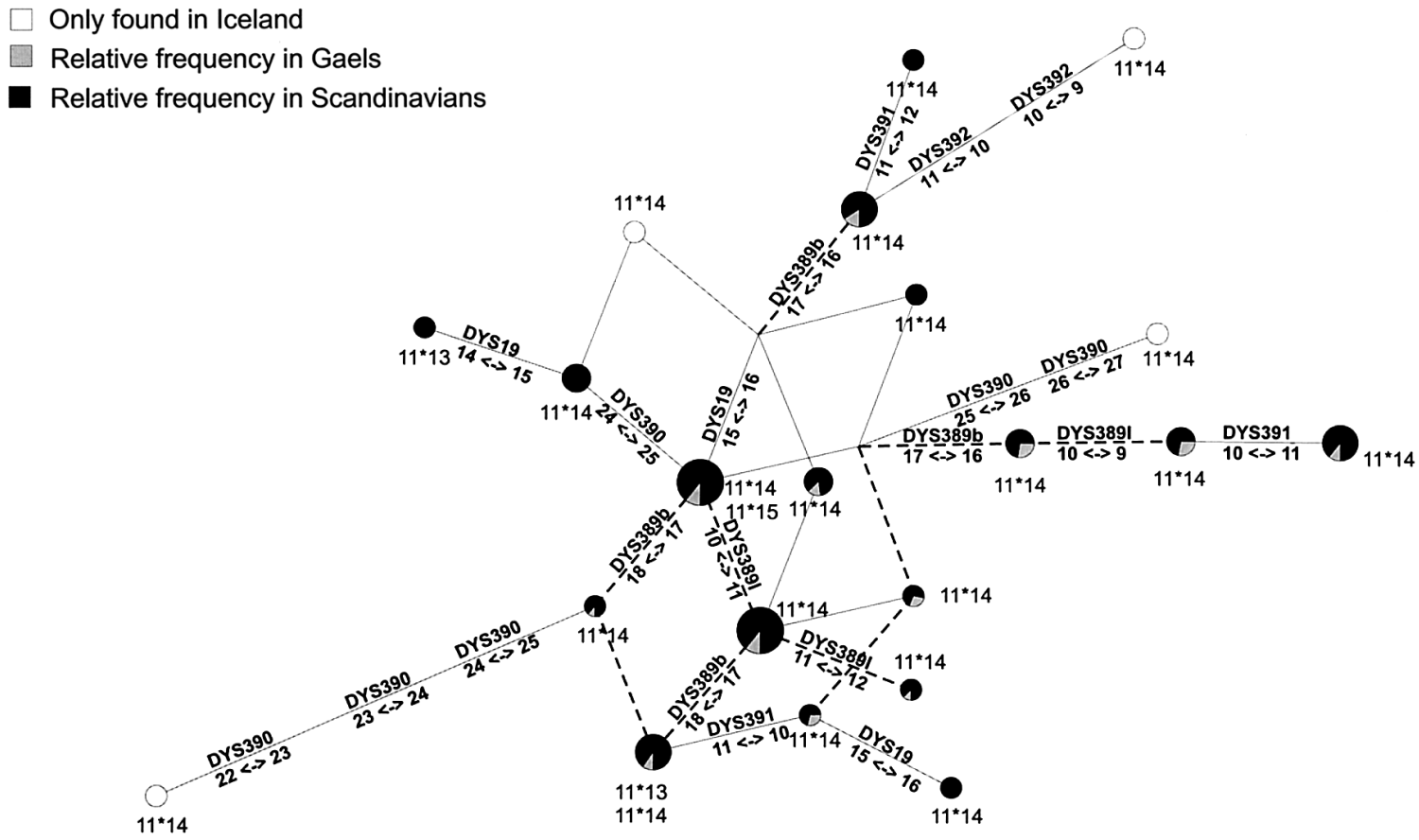

Figure 5 Median-joining network showing the phylogenetic relationships of Icelandic Y-chromosome haplotypes within haplogroup 3. For definitions, see the legend for figure 3.

in this study. Although their divergence is appropriately represented in the network, the evolutionary pathway connecting these particular haplotypes to the rest of haplogroup 2 must be viewed as tentative. Notwithstanding this uncertainty, the pattern of haplotype sharing between populations reveals that, as with haplogroup 1, most of the haplogroup-2 chromosomes are found in both Scandinavians and Gaels, although seven chromosomes are shared exclusively by Icelanders and Scandinavians. Only one haplotype occurs at a higher frequency in Gaels than in Scandinavians. That it is one of the divergent haplotypes lends support to the idea that these haplotypes do not share a recent monophyletic origin with the rest of the haplogroup-2 chromosomes. Haplogroup 3 (fig. 5) presents a similar pattern of haplotype sharing, with a clear Scandinavian prevalence. Here we observe four haplotypes shared with both Scandinavians and Gaels and four haplotypes shared exclusively with Scandinavians (when mutations at DYS388, DYS389I, and DYS389b are excluded). For both haplogroups 2 and 3, DYS385 allele sizes are consistent with the reconstructed phylogenies.

Summary Statistics and Genetic Distances

A summary of the genetic diversity present within populations and haplogroups is shown in table 4. In accordance with expectations based on historical population sizes, the Icelanders have low gene diversity and, judging from θ_k values, a relatively small effective population size of males (N_{me}). However, when diversity is measured in terms of the ASDs between haplotypes within a population, the Scots and Irish are shown to have, by far, the least-diverse collection of haplotypes, followed by the Icelanders. The much smaller within-population ASD for the Gaels is a direct result of the fact that their Y chromosomes predominantly belong to haplogroups 1 and 2. The coexistence, in a population, of chromosomes belonging to haplogroups 2 and 3 (which is more typical in Scandinavians) results in a much larger ASD between haplotypes, because these haplogroups are separated by three diallelic mutations (see fig. 1).

If ASDs between haplotypes are examined within haplogroups, it emerges that the Icelanders have a slightly higher ASD for haplogroup-1 chromosomes than do any of the source populations, and they have a higher ASD for haplogroup-2 chromosomes than do the Scandinavians. This is consistent with an admixture scenario in which the Icelanders have received, from both source populations, chromosomes belonging to both haplogroups and in which the source populations are associated with different subclades within haplogroups. This is particularly evident for haplogroup 1, in which the largest mutational differences are between the exclusively Scandinavian "branch-A" haplotypes and the rest of the haplotypes, most of which are either predominantly or exclusively found in the Gaels (see fig. 3). The relatively high percentage of private haplotypes belonging to haplogroup 1 in Scandinavia indicates that these chromosomes could have a fairly long history in the Norse population. Thus, they are unlikely to be the result of gene flow from the British Isles—for example, as a result of the import of Gaelic slaves during the Viking period. Similarly, the high percentage of private haplotypes belonging to haplogroup 2 in the Gaels suggests that these are likely to be only partially derived from Viking activities in Ireland and Scotland. In the case of haplogroup 3, the Icelanders exhibit less diversity than do the Scandinavians. Because only five haplogroup-3 chromosomes are observed in the Gaels and because the Icelanders exhibit less haplogroup-3 diversity than do the Scandinavians, it seems reasonable to conclude that Icelandic haplogroup-3 chromosomes all originated from Scandinavia.

The Icelanders have the densest sampling saturation and a low proportion of private haplotypes. This is likely to be a reflection of both the admixed nature of the Icelandic Y-chromosome pool and its relative scarcity of haplotypes (reflected in the low θ_k value). The high proportion of private haplotypes in the Irish is primarily due to a large number of haplogroup-1 lineages that are not found in any of the other populations. This is likely to be due to a combination of dense sampling saturation from the Irish Y-chromosome pool (see the "Sample Saturation Ratio" column in table 4) and divergence from the Y-chromosome pools of the other populations. The Scots have a much lower proportion of private haplotypes, but they also have a lower relative sampling saturation. Both the Norwegians and the Swedes also have relatively low sampling saturation, and the very small sample size of the Danes precludes any meaningful inferences about this population.

It is possible to estimate coalescence dates for the chromosomes of haplogroups 1, 2, and 3, because they are well represented in the Icelandic, Scandinavian, and Gaelic samples. Taking the modal haplotype within each of these haplogroups as the ancestral haplotype and using the restricted set of microsatellite loci (DYS19 and DYS390–DYS393), we estimated the ASD between the ancestral haplotype and all other haplotypes. These ASD values, per locus, for haplogroups 1, 2, and 3 were 0.612, 0.745 and 0.448, respectively. Given a paternal generation time of 35 years (Tremblay and Vézina 2000) and an average mutation rate of 0.0021 per locus per generation (95% CI 0.0006–0.0049) (Heyer et al. 1997), these figures correspond to dates of 10,200 (4,371–35,700), 12,497 (5,321–43,458), and 7,467 (3,200–26,133) years before the present (YBP) for haplogroups 1, 2, and 3, respectively.

A summary of the overall relationship between the

Table 4

Summary Statistics for Y-Chromosome Haplotypes

								Haplogroup 1			Haplogroup 2			Haplogroup 3			Haplogroup 16		
Population[a]	Sample Size	k	Private K[b] (%)	Gene Diversity	θ_k	Sample Saturation Ratio[c]	Total ASD[d]	ASD	k	Private k (%)	ASD	k	Private k (%)	ASD	k	Private k (%)	ASD	k	Private k (%)
Irish	222	66	50	.94	31.39	.77	138.21	6.23	39	56.4	11.52	19	31.6	0	1	.0	0	1	100.0
Scots	61	26	23.1	.916	16.60	.38	160.29	4.8	17	29.4	4.59	4	25.0	2.87	4	.0	NA	0	NA
Norwegians	112	48	43.7	.943	31.29	.39	336.43	6.52	18	38.9	4.94	11	45.4	4.57	12	50.0	.44	1	.0
Swedes	109	45	31.1	.957	28.18	.42	340.85	5.8	9	22.2	4.31	19	15.8	5.95	11	36.4	1.47	4	50.0
Icelanders	181	50	28	.921	22.50	.81	286.57	7.42	21	28.6	5.61	17	17.6	2.99	11	36.4	0	1	100.0
Scandinavians	233	76	48.8	.966	44.63	.57	339.98	5.86	22	40.9	5.32	29	34.5	5.64	20	60.0	1.22	4	100.0
Gaels	283	82	53.9	.947	33.74	.91	143.24	6.08	45	64.4	10.34	20	35.0	2.72	4	.0	0	1	100.0

Note.—Statistics for Y-chromosome haplotypes, based on the common set of loci (92R7, M9, SRY-1532, YAP, TAT, SRY-2627, DYS19, and DYS390–DYS393) for all populations.

[a] Scandinavians represent a combined sample of 112 Norwegians, 109 Swedes, and 12 Danes. Summary statistics for Danes are not shown, because the sample size was too small to provide reliable estimates. Gaels represent a combined sample of 222 Irish and 61 Scots.

[b] Percentage of haplotypes from a population that are not found in the other populations (including the 12 Danish samples).

[c] Relative sampling saturation (see the Material and Methods section). GD = gene diversity.

[d] Where differences at diallelic loci are weighted 1,000 times higher than differences at microsatellite loci, which are given a value of 1. Numbers are presented as multiples of 10^4.

Table 5

ASD Values within and between Populations, by Haplogroups and by Weighted Averages Across Haplogroups

	ASD[a], within and between Populations, for									Weighted-Average ASD[b]		
	Haplogroup 1 (k = 61)			Haplogroup 2 (k = 39)			Haplogroup 3 (k = 24)			Haplogroups 1–3		
Population	Gae	Ice	Sca	Gae	Ice	Sca	Gae	Ice	Sca	Gae	Ice	Sca
Gae	6.08	.76	.45	10.34	.44	1.16	2.72	.51	.14		.61	.61
Ice	7.51	7.42	−.10	8.41	5.61	.52	3.37	2.99	.14	6.99		.14
Sca	6.42	6.54	5.86	8.99	5.98	5.32	4.32	4.45	5.64	6.82	5.96	

Note.—Gae = Gaels; Ice = Icelanders; Sca = Scandinavians.

[a] For each haplogroup, the lower diagonal shows ASDs between all pairs of haplotypes belonging to a particular haplogroup in two populations. ASDs of all pairs of haplotypes within populations are shown (*underlined*) in the diagonal. The upper diagonal contains ASD distances corrected for within-population diversity; the average within-population diversity of the two populations being compared has been subtracted from the uncorrected ASD distances presented in the lower diagonal.

[b] Weighted by the total number of haplotypes belonging to each haplogroup. The lower diagonal contains weighted average for uncorrected ASD values, whereas the upper diagonal contains the weighted average for corrected ASD values.

haplotypes of the Icelanders and those found in Scandinavians and Gaels can be obtained by calculation of genetic distances between these populations. Table 5 shows genetic distances, based on ASD, between these three populations, for the three main haplogroups, and the weighted averages of these distances across haplogroups. The uncorrected ASD values (lower diagonals in table 5) indicate a closer relationship between Icelandic and Scandinavian haplotypes than between Icelandic and Gaelic haplotypes for haplogroups 1 and 2, but not for haplogroup 3. When these values are corrected for within-population diversities (corrected ASD values are in the upper diagonals in table 5), a closer relationship between Icelandic and Scandinavian haplotypes is revealed for haplogroups 1 and 3 but not for haplogroup 2. The summary of these results provided by the weighted averages of ASD values (both corrected and uncorrected) across haplogroups clearly indicates a closer relationship between the Icelandic and Scandinavian Y-chromosome pools than between the Icelanders and Gaels.

Genetic distances between the populations, based on haplotype frequency differences, are shown in the lower diagonal of table 6. Here again we observe a closer relationship between the Icelandic and Scandinavian Y-chromosome pools. The upper diagonal of table 6 presents the proportion of all haplotypes from two populations that are shared, demonstrating the greater overlap between the Y-chromosome pools of the Icelanders and Scandinavians.

The population relationships revealed by the ASD values, chord distances, and lineage sharing are supported by genetic distances based on ρ. Thus, the ASD between Icelandic haplotypes and their closest counterparts in Scandinavians (0.19) is considerably smaller than the same distance between Icelandic haplotypes and those of the Gaels (0.35). As would be expected of a population with ancestral contributions from both Scandinavians and Gaels, ρ is reduced to 0.163 when the Icelanders are compared with a combined sample of these putative parental populations. This ρ value can be used to provide an estimate of the time-depth of the mutational divergence of Icelandic haplotypes from the two source Y-chromosome pools (see Forster et al. 1996). Using a generation time of 35 years (Tremblay and Vézina 2000) and an average mutation rate of 0.0021 per locus per generation (95% CI 0.0006–0.0049) (Heyer et al. 1997), we obtain a divergence date of 2,717 YBP (1,164–9,508). There is no archaeological or historical evidence to support such an ancient settlement of Iceland. It is likely that this discrepancy results from the fact that many of the Icelandic Y-chromosome haplotypes currently identified as private have yet to be sampled from the Scandinavian and Gaelic source Y-chromosome pools.

Table 6

Haplotype Divergence between Populations and Lineage Sharing

	Gaels	Icelanders	Scandinavians
Gaels		.247	.244
Icelanders	.0113		.32
Scandinavians	.0125	.0086	

Note.—The lower diagonal shows haplotype divergence (on the basis of the chord distance) between populations, calculated using haplotype frequencies. The upper diagonal shows the proportion of all haplotypes from two populations that are present in both populations.

Admixture Estimates

Although the summary statistics above clearly identify Scandinavia as the major source of Icelandic Y chromosomes, the evidence clearly suggests a significant Gaelic contribution. Here we estimate the relative proportions of Scandinavian and Gaelic ancestry of Icelandic Y chromosomes, using the three admixture methods (m_R, m_Y, and m_r) outlined in the Material and Methods section. It is useful to start by examining a scatterplot of the haplotype frequency differentials that lie at the heart of the standard least-squares admixture estimator.

Figure 6 shows the relationship between $p_I - p_G$ and $p_S - p_G$, where p_I, p_G, and p_S represent haplotype frequencies in the Icelanders, Gaels, and Scandinavians, respectively. Haplotypes are defined by the same set of loci used to compare these populations in previous sections. The least-squares regression line is shown, and its slope (0.758) approximates the m_R estimate of Scandinavian ancestry of Icelandic Y chromosomes. However, the residuals from the regression line are relatively large and indicate the extent to which haplotype frequencies in the Icelanders, Scandinavians, and Gaels have been affected by evolutionary change (i.e., genetic drift, selection, or migration) and/or the extent of error caused by sampling.

Table 7 presents the admixture estimates derived from the three different methods. Interestingly, m_R and m_Y provide quite divergent estimates of Gaelic ancestry in the Icelandic Y-chromosome pool (25.6% and 40.7%, respectively). If the heuristic method m_ρ is used to assess the goodness of fit of admixture models to the observed haplotype distribution in the admixed and source populations, we obtain a best estimate of 19.5% Gaelic admixture. This is closer to the m_R estimate, which falls well within the 95% CI of 0.045–0.365 of Gaelic ancestry for m_ρ. Judging from this, the considerably larger m_Y value seems likely to be an overestimate. To understand the reason for this difference, it is informative to compare results based only on haplogroup frequencies as defined by the five diallelic loci. For this data set of diallelic haplotypes, the m_Y estimate of Scandinavian ancestry remains essentially unchanged at 0.594, whereas both m_R and m_ρ are reduced, in a similar way, to 0.671 and 0.72, respectively. The reduction in the latter two indices is caused by a few cases of disproportionate haplotype sharing between Icelanders and Scandinavians; such cases are revealed only when microsatellite loci are included. The m_Y estimate is, however, strongly constrained by the mutational differences that define the phylogenetic tree of diallelic haplotypes (see fig. 2). This emerges clearly when m_Y is calculated for the haplotypes defined by diallelic loci and when the distances between all pairs of haplotypes are set to the same value. When the underlying phylogeny of the diallelic loci is ignored in this way, m_Y rises to 0.695. Es-

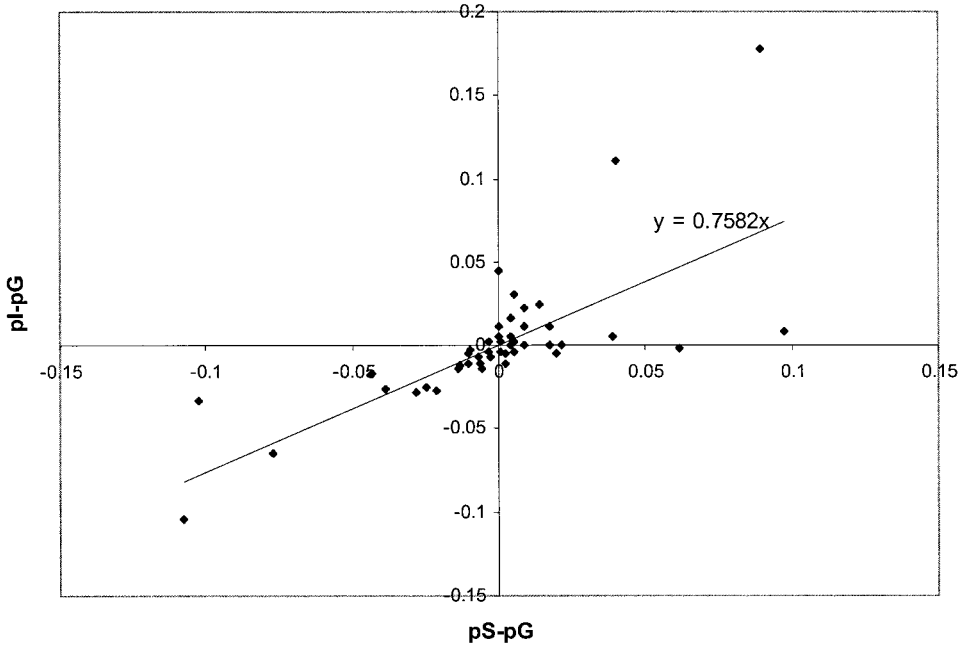

Figure 6 The scatterplot presents the relationship between $p_I - p_G$ and $p_S - p_G$, where p_I, p_G, and p_S denote the haplotype frequencies for the Icelanders, Gaels, and Scandinavians, respectively. The slope of the least-squares fitted regression line approximates the m_R estimate of Scandinavian ancestry in Icelandic Y chromosomes.

Table 7

Estimates of Scandinavian and Gaelic Ancestry in the Icelandic Y-Chromosome Pool

Parental Population	m_R	m_Y	m_ρ	1,000 Bootstrap Iterations		Goodness of Fit
				m_R (SD)	m_Y (SD)	m_ρ 95% CI
Scandinavians	.744	.593	.805	.721 (.088)	.597 (.072)	.715–.875
Gaels	.256	.407	.195	.279 (.088)	.403 (.0712)	.125–.285

timates based on m_Y are most effective when genetic differences between source populations are due to accumulated mutations over a long period of time, but they are less effective when these differences are due to genetic drift or when there has been independent gene flow between source populations. Clearly, the divergence time between the Scandinavians and the Gaels is insufficient for m_Y to provide a valid estimate, and it does not help that there may have been gene flow between these two source populations of the Icelandic Y-chromosome pool. If we are correct in placing more confidence in estimates provided by m_R and m_ρ, then the proportion of Gaelic males in the Icelandic founding population could have been somewhere between 19.5% and 25.6%.

Discussion

Our findings show that most Icelandic Y chromosomes are tightly clustered with regard to variation at microsatellite loci within the three main haplogroups. However, a few haplogroup-1 and -2 chromosomes show marked divergence from these tight clusters, such that they cannot be said to form a meaningful phylogenetic unit with the rest. Haplogroups 1 and 2 both lie at central positions in the Y-chromosome phylogeny (see fig. 1) and are thus more likely to include a fusion of deeply divergent haplotype clusters, which are not differentiated by the diallelic loci available for this study. This is less of a problem for haplogroup 3 (a relatively young subcluster of haplogroup 1 defined by a mutation at SRY-1532), whose haplotypes form a more obvious monophyletic group. The ability to detect subclusters of haplogroups 1 and 2 was due to the combination of diallelic loci and a relatively large number of microsatellites. We demonstrate, for example, that the DYS385 microsatellite can be usefully recruited for phylogeographic studies of the Y chromosome.

The results of this study indicate that most of the Gaelic males in the Icelandic founding population are likely to have carried haplogroup-1 Y chromosomes, but a slight contribution of haplogroup-2 chromosomes from the British Isles cannot be ruled out. Hill et al. (2000) have proposed that haplogroup-1 chromosomes were indigenous to the early population of Ireland but that Irish chromosomes belonging to other haplogroups could be associated either with prehistoric or more-recent historical immigration to the island. On the basis of the relative frequencies, in Scandinavians and Gaels, of some haplogroup-2 chromosomes shared with Icelanders, it seems likely that some of them may predate the Viking period. The same applies to some of the haplogroup-1 chromosomes currently found in Scandinavia (branch A in fig. 3), which may turn out to constitute a separate haplogroup.

The relative ASD diversity of Icelandic haplotypes currently assigned to haplogroups 1 and 2 suggests a dual origin from both Scandinavia and the British Isles. In contrast, the relatively low diversity of haplogroup-3 chromosomes in the Icelanders and their scarcity in the Gaelic populations point toward a single source in Scandinavia. This lower diversity is a reflection of the extent to which the Icelandic Y-chromosome pool has been shaped by the initial genetic sampling event of the settlement and the subsequent effects of genetic drift. This aspect of Icelandic population history is also revealed in a relatively low θ_k value, reflecting a small effective population size of males.

The Y chromosome is just one locus and, thus, it can provide only a partial view of the overall ancestry of the Icelandic gene pool. However, the information gained from this locus is unique in that it traces the patrilineal ancestors of contemporary Icelandic males. We used three different admixture estimators in an attempt to reveal the provenance of these ancestors. The m_Y estimator (based on the frequencies and mean coalescent times of haplotypes within and between populations) indicates a relatively high estimate (40.7%) of Gaelic ancestry. However, this result seems to be strongly influenced by phylogenetic relationships of the ancient diallelic mutations and their frequencies in the Scandinavian and Gaelic source populations. Because these source Y-chromosome pools have been shaped more by gene flow and/or genetic drift than by the mutational divergence of diallelic haplotypes, the coalescence times of the diallelic haplotypes are less informative for admixture estimates than their frequencies. However, results based on m_R (a least-squares estimator based solely on observed founder haplotype frequencies) and m_ρ (a new goodness-of-fit estimator based on

the frequencies of observed and inferred founder haplotypes) are broadly consistent, respectively indicating 25.6% and 19.5% Gaelic ancestry of Icelandic Y chromosomes. Inferences about the remaining half of the founders—the females—can be made using the matrilineal inheritance pattern of mtDNA. The results of a recent analysis of mtDNA variation in Iceland suggest that the majority of Icelandic founding females may have originated from the British Isles (Helgason et al. 2000). Therefore, on the basis of evidence from these two uniparental markers, it is possible to conclude that the Icelandic founding population was comprised of a majority of males carrying Scandinavian Y chromosomes and females carrying Gaelic mtDNA lineages. In addition to purely Scandinavian males and Gaelic females, many of these Icelandic founders are likely to have been the offspring of admixed family groups formed by Scandinavian males and Gaelic females in the British Isles prior to the settlement of Iceland.

As with any admixture analysis, the degree of confidence that we can place in our estimates of Norse and Gaelic ancestry of Icelandic Y chromosomes is dependent on the assumption that the gene pools of the populations in Iceland, Scandinavia, and the British Isles have not been drastically altered by genetic drift during the past 1,100 years. There are at least two reasons to believe that haplotype frequencies in the Scandinavian and Gaelic source Y-chromosome pools have remained relatively stable. First, as is apparent from figure 2, haplogroup frequencies within the British Isles (in Ireland, Scotland, and England) are almost identical, and the same is true for haplogroup frequencies within Scandinavia (Norway and Sweden). Second, Y-chromosome haplotype and allele frequencies from both groups contribute to European-wide clinal patterns, generated by population processes much older than the settlement of Iceland (Zerjal et al. 1997; Casalotti et al. 1999; Lucotte and Loirat 1999; Zerjal et al. 1999; Hill et al. 2000). Neither observation is consistent with substantial random fluctuation of haplotype frequencies within these populations resulting from drift during the past 1,100 years.

Greater random fluctuations in haplotype frequencies would be expected in the Icelanders, given their smaller population size, geographic isolation, and history of population crashes. Indeed, this is indicated by the residuals from the fitted regression line of the points $p_I - p_G$ and $p_S - p_G$ in figure 6, which reflect a sizeable variance in admixture estimates for individual lineages ($\sigma^2 = 1.85$). The effects of drift are also manifest in the relatively wide 95% CIs for m_ρ (CI 0.715–0.875) and m_R (CI 0.549–0.893; estimated from table 7). Nonetheless, even these ranges indicate that the majority of Icelandic founding males originated from Scandinavia.

More knowledge about the history and geographic spread of Y chromosomes in Europe will resolve uncertainties resulting from the use of modern populations as indicators of ancestral gene pools. In a recent study, for example, Zerjal et al. (1999) proposed an Asian origin for haplogroup-3 chromosomes and demonstrated a cline of decreasing frequency from Asia into Europe, with high frequencies in northeast Europe and very low frequencies in the British Isles and southwest Europe. Using additional phylogenetic information, Zerjal et al. (1999) suggested that this distribution is the result of a population spread from Asia into Europe that did not reach these latter regions. This clearly supports the idea that Icelandic haplogroup-3 chromosomes arrived with males of Scandinavian origin and that Gaelic chromosomes belonging to this haplogroup may be a result of Viking activities in the British Isles.

The various analyses presented in this paper, whether of genetic distances, lineage sharing, or admixture estimates, are consistent in their indication of a close relationship between the contemporary Icelandic and Scandinavian Y-chromosome pools. Given that the patrilineal founding ancestors of contemporary Icelandic males are equivalent, in terms of their ancestry, to the many founding males who did not leave patrilineal descendants in the contemporary population and given that, at present, there is no reason to suppose otherwise, then it seems reasonable to conclude that the large majority of founding males carried Scandinavian Y chromosomes.

Acknowledgments

We would like to thank the volunteers who contributed their DNA for this study. We are indebted to Tatiana Zerjal, Lars Beckman, and Chris Tyler-Smith, for information about unpublished Swedish Y-chromosome data and for useful comments on the manuscript. We thank two anonymous reviewers for their perceptive comments. The typing of the Scottish and Norwegian data was supported by a Wellcome Trust award (to B.S.). A.H. was supported by an Overseas Research Studentship award from the Committee of Vice-Chancellors and Principals of Universities and Colleges in the United Kingdom, 1997–99.

Electronic-Database Information

The URLs for data in this article are as follows:

ADMIX: Inferring Admixture Proportion from Molecular Data, http://www.unife.it/genetica/Giorgio/giorgio_soft.html#ADMIX (for ADMIX1_0 software)

Life Sciences and Engineering Technology Solutions, http://www.fluxus-engineering.com/ (for Network 2.0 software)

Y-STR Haplotype Reference Database, http://ystr.charite.de/index_gr.html

References

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16: 37–48

Bernstein F (1931) Die geographische Verteilung der Blut-gruppen und ihre anthropologische Bedeutung. In: Comitato Italiano per o studio dei problemi della populazione. Instituto Poligrafico dello Stato, Rome, pp 227–243

Bertorelle G, Excoffier L (1998) Inferring admixture proportions from molecular data. Mol Biol Evol 15:1298–1311

Bjarnason Ó, Bjarnason V, Edwards JH, Friðriksson S, Magnússon M, Mourant AE, Tillis D (1973) The blood groups of Icelanders. Ann Hum Genet 36:425–458

Calafell F, Pérez-Lezaun A, Bertranpetit J (2000) Genetic distances and microsatellite diversification in humans. Hum Genet 106:133–134

Casalotti R, Simoni L, Belledi M, Barbujani G (1999) Y-chromosome polymorphisms and the origins of the European gene pool. Proc R Soc Lond B Biol Sci 266:1959–1965

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. Am J Hum Genet 19:233–257

Clover CJ (1988) The politics of scarcity: notes on the sex ratio in early Scandinavia. Scandinavian Studies 60:147–188

Davies N (1999) The Isles: a history. Macmillan, London

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112

Forster P, Harding R, Torroni A, Bandelt HJ (1996) Origin and evolution of native American mtDNA variation: a reappraisal. Am J Hum Genet 59:935–945

Forster P, Röhl A, Lünnemann P, Brinkmann C, Zerjal T, Tyler-Smith C, Brinkmann B (2000) A short tandem repeat–based phylogeny for the human Y chromosome. Am J Hum Genet 67:182–196

Guldberg P, Zschocke J, Dagbjartsson A, Henriksen KF, Guttler F (1997) A molecular survey of phenylketonuria in Iceland: identification of a founding mutation and evidence of predominant Norse settlement. Eur J Hum Genet 5:376–381

Hammer MF, Horai S (1995) Y-chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Helgason A, Sigurðardóttir S, Gulcher JR, Ward R, Stefánsson K (2000) mtDNA and the origin of the Icelanders: deciphering signals of recent population history. Am J Hum Genet 66:999–1016

——. Sampling saturation and the European mtDNA pool: implications for detecting genetic relationships among populations. In: Renfrew C, Boyle K (eds) Molecular genetics and early Europe: working papers in population prehistory. The McDonald Institute, Cambridge (in press)

Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997) Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. Hum Mol Genet 6:799–803

Hill EW, Jobling MA, Bradley DG (2000) Y-chromosome variation and Irish origins. Nature 404:351–352

Hurles ME, Irven C, Nicholson J, Taylor PG, Santos FR, Loughlin J, Jobling MA, Sykes BC (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. Am J Hum Genet 63: 1793–1806

Hurles ME, Veitia R, Arroyo E, Armenteros M, Bertranpetit J, Perez-Lezaun A, Bosch E, Shlumukova M, Cambon-Thomsen A, McElreavey K, Lopez de Munain L, Rohl A, Wilson IJ, Singh LJ, Pandya A, Santos FR, Tyler-Smith C, Jobling MA (1999) Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. Am J Hum Genet 65:1437–1448

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124

Jones G (1984) A history of the Vikings. Oxford University Press, Oxford

Karafet TM, Zegura SL, Posukh O, Osipova L, Bergen A, Long J, Goldman D, Klitz W, Harihara S, de Knijff P, Wiebe V, Griffiths RC, Templeton AR, Hammer MF (1999) Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. Am J Hum Genet 64:817–831

Karras RM (1988) Slavery and society in medieval Scandinavia. Yale University Press, New Haven

Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. Am J Hum Genet 66:1580–1588

Lucotte G, Loirat F (1999) Y-chromosome DNA haplotype 15 in Europe. Hum Biol 71:431–437

Pálsson B, Pálsson F, Perlin M, Guðbjartsson H, Stefánsson K, Gulcher J (1999) Using quality measures to facilitate allele calling in high-throughput genotyping. Genome Res 9: 1002–1012

Pálsson H, Edwards P (trans) (1972) The book of settlements: landnámabók. University of Manitoba, Winnipeg

Pfeiffer H, Brinkmann B, Huhne J, Rolf B, Morris AA, Steighner R, Holland MM, Forster P (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. Int J Legal Med 112:291–298

Previdere C, Stuppia L, Gatta V, Fattorini P, Palka G, Tyler-Smith C (1999) Y-chromosomal DNA haplotype differences in control and infertile Italian subpopulations. Eur J Hum Genet 7:733–736

Roberts DF, Hiorns R (1965) Methods of analysis of the genetic composition of a hybrid population. Hum Biol 37: 38–43

Santos FR, Pandya A, Tyler-Smith C, Pena SDJ, Schanfield M, Leonard WR, Osipova L, Crawford MH, Mitchell RJ (1999) The central Siberian origin for Native American Y chromosomes. Am J Hum Genet 64:619–628

Sawyer P (1997) The age of the Vikings and before. In: Sawyer P (ed) The Oxford illustrated history of the Vikings. Oxford University Press, Oxford, pp 1–18

Sayers W (1994) Management of the Celtic fact in Landnámabók. Scandinavian Studies 66:129–153

Sigurðsson G (1988) Gaelic influence in Iceland. Menningarsjóður, Reykjavík

Steffensen J (1975) Menning og meinsemdir: ritgerðarsafn um mótunarsögu íslenskrar pjóðar og baráttu hennar við hungur og sóttir. Ísafoldarprentsmiðja, Reykjavík

Tremblay M, Vézina H (2000) New estimates of intergenerational time intervals for the calculation of age and origins of mutations. Am J Hum Genet 66:651–658

Thompson EA (1973) The Icelandic admixture problem. Ann Hum Genet 37:69–80

Tyler-Smith C (1999) Y-chromosomal DNA markers. In: Papiha S, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Kluwer Academic, New York, pp 65–73

Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 7:996–1005

Wijsman EM (1984) Techniques for estimating genetic admixture and applications to the problem of the origin of the Icelanders and the Ashkenazi Jews. Hum Genet 67:441–448

Williams JT (1993) Origin and population structure of the Icelanders. Hum Biol 65:167–191

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhovel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60: 1174–1183

Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgrafov O, Singh L, Thangaraj K, Destro-Bisol G, Thomas MG, Qamar R, Mehdi SQ, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C (1999) The use of Y-chromosomal DNA variation to investigate population history. In: Papiha S, Deka R, Chakraborty R (eds) Genomic diversity: applications in human population genetics. Plenum, New York, pp 91–101