Am. J. Hum. Genet. 62:1171–1179, 1998

Dual Origins of Finns Revealed by Y Chromosome Haplotype Variation

Rick A. Kittles,[1] Markus Perola,[4] Leena Peltonen,[4] Andrew W. Bergen,[2] Richard A. Aragon,[2] Matti Virkkunen,[5] Markku Linnoila,[3] David Goldman,[2] and Jeffrey C. Long[1]

[1]Section on Population Genetics and Linkage, [2]Laboratory of Neurogenetics, and [3]Laboratory of Clinical Studies, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda; [4]Department of Human Molecular Genetics, National Public Health Institute, and [5]Department of Psychiatry, University of Helsinki, Helsinki

Summary

The Finnish population has often been viewed as an isolate founded 2,000 years ago via a route across the Gulf of Finland. The founding event has been characterized as involving a limited number of homogeneous founders, isolation, and subsequent rapid population growth. Despite the purported isolation of the population, levels of gene diversity for the Finns at autosomal and mitochondrial DNA loci are indistinguishable from those of other Europeans. Thus, mixed or dual origins for the Finns have been proposed. Here we present genetic evidence for the dual origins of Finns by evaluating the pattern of Y chromosome variation in 280 unrelated males from nine Finnish provinces. Phylogenetic analysis of 77 haplotype configurations revealed two major star-shaped clusters of Y haplotypes, indicative of a population expansion from two common Y haplotypes. Dramatic and quite significant differences in Y haplotype variation were observed between eastern and western regions of Finland, revealing contributions from different paternal types. The geographic distribution and time of expansion for the two common Y haplotypes correlate well with archeological evidence for two culturally and geographically distinct groups of settlers. Also, a northeastern to southwestern gradient of Y haplotype frequencies provides convincing evidence for recent male migration from rural areas into urban Finland.

Introduction

After the emergence of the "Finnish disease heritage" concept (Norio et al. 1973) and recent advances in molecular genetics, interest in the Finnish population, for studies on genetic disorders, has increased rapidly. At present, there are >33 rare genetic diseases that are more prevalent in Finland than in other populations (de la Chapelle 1993). The high prevalence of these rare diseases has been attributed to founder effects resulting from Finland's unique population history. This uniqueness is conspicuously reflected by the fact that, unlike most other Europeans, Finns do not speak an Indo-European language. Finnish is a member dialect of the Uralic language family, whose only other speakers within Europe are found in Estonia, Hungary, and the circumpolar regions of Finland, Sweden, Norway, and Russia (Saami) (Ruhlen 1987).

Two opposing theories on the origin of the Finnish population exist. The single-origin model purports a founding of Finland ~2,000 years ago by a small number of settlers (Nevanlinna 1972; de la Chapelle 1993), followed by relative isolation. Internal migration was minimal until the 17th century, when population movements into late settlements of the north and east of Finland began (Norio 1981; de la Chapelle 1993). During the period of isolation, population growth was not constant because of wars, famine, and disease (Peltonen et al. 1995), and it was not until the last 300 years that a population expansion occurred—from ~250,000 individuals to >5,000,000. A competing model, the dual-origins hypothesis, contends that two different groups settled Finland. The first group arrived from the east, near the Lake Ladoga region, and the second group arrived from the south via the Gulf of Finland (Eriksson 1973; Meinander 1973; Norio 1981). Like its opposing model, the dual-origin model asserts that the Finnish population has remained isolated for ~2,000 years, especially in some of the more sparsely populated regions of the country. Consequently, the stochastic effects of genetic drift over ~80 generations greatly shaped the geographic distribution of disease alleles.

The close linguistic affinity between Finns and Estonians has been the basis for those who trace a single origin of the Finnish population to an area south of the Gulf of Finland that includes present-day Estonia (Nevanlinna 1972; Luho 1976; Fodor and Czeizel 1991). However, archeological evidence supports the dual-ori-

Received December 5, 1997; accepted for publication March 13, 1998; electronically published April 17, 1998.

Address for correspondence and reprints: Dr. Rick Kittles, Howard University Cancer Center, 2041 Georgia Avenue, Room 507, NW, Washington, DC 20060. E-mail: rkittles@howard.edu

0002-9297/98/6205-0023$02.00

gins model, with considerable cultural differences between eastern and western Finland. The cultural differences, which include eastern and western varieties of plows, sleighs, and architecture (Luho 1976; Vilkuna 1976), date back almost 1,000 years (Hajdu 1975).

Genetic evidence for the single-origin model was claimed in studies of blood antigens (Nevanlinna 1972, 1973). However, these marker systems possess low resolving power for assessing population structure. Reduced gene diversity has yet to be reasonably demonstrated in autosomal and mitochondrial DNA studies of the Finns (Kittles et al. 1996; Lahermo et al. 1996). However, genetic evidence for a Finnish population bottleneck was recently reported as a marked reduction in Y chromosome diversity for Finns as compared with other populations (Sajantila et al. 1996). Although reduced levels of Y chromosome haplotype diversity may appear to support a single origin of the Finns, it reveals little, if anything, about the different types of Y chromosomes in the Finnish population or about their ancestry.

Here we determine whether the single- or dual-origin model best explains the diversity, distribution, and lineages of Y chromosomes in the Finnish population. Compound haplotypes were constructed by use of seven Y-specific microsatellite loci, a restriction site at the Y alphoid satellite *DYZ3* locus (Santos et al. 1995), and a deletion polymorphism at the *DYF155S2* locus (Jobling et al. 1996) for 280 unrelated males from nine Finnish provinces. The deletion polymorphism is common in populations with Asian ancestry and has been observed in Finland at a frequency of ~55% (Jobling et al. 1996). Y chromosome compound haplotypes constructed by use of these Y-specific loci are extremely informative for population genetic studies, mainly because of the high mutation rates of microsatellite loci, their paternal/clonal inheritance, and an effective population size, ~25% that of autosomes. Here we present genetic evidence in support of two independent groups of settlers founding Finland, one from Asia and the other from Europe.

Material and Methods

DNA Samples

DNA was obtained from 280 unrelated men born in Finland. All subjects had Finnish surnames and were participants in one of two long-term genetic studies of the Finnish population: a National Public Health Institute (Finland) project on cardiovascular disease risk factors (Vartiainen et al. 1994) and a National Institute on Alcohol Abuse and Alcoholism, (U.S. NIH) project on the genetics of alcoholism and related behaviors (Virkkunen et al. 1994).

Polymorphism Typing

Genotyping was performed blind to geographic origin of samples. Primer sequences for two trinucleotide repeat loci, *DYS388* and *DYS392*, and five tetranucleotide repeat loci *DYS389*, *DYS390*, *DYS391*, *DYS393*, and *DYS394*, were identified from the Genome Data Base and tested for Y chromosome specificity. *DYS394* is the same locus as the commonly used *DYS19* locus. Primers for the *DYS389* locus amplified two products: a smaller (240–260 bp) and a larger (370–400 bp) product. Only the smaller product, designated *DYS389a*, was used for the analyses. For each PCR, 50 ng of genomic DNA was added to 200 uM of dNTPs, 10 mM Tris-HCl (pH = 8.3), 50 mM KCl, 1.0–2.0 mM $MgCl_2$, 0.6 U AmpliTaq polymerase (Perkin Elmer), and 0.33 μM of primers. The forward primers were fluorescently labeled. The PCR cycling conditions consisted of 93°C for 3 min; 10 cycles at 94°C for 15 s, 55°C for 15 s, and 72°C for 30 s; 20 cycles at 89°C for 15 s, 55°C for 15 s, and 72°C for 30 s; and an extension cycle at 72°C for 10 min. The PCR products were then pooled in the presence of a size standard (Genescan 500 GS) and electrophoresed by use of an ABI 373A DNA sequencer. Alleles at each locus were determined by the GS Analysis and Genotyper programs (ABI). Corrections were made for individual gel shifts by means of the computer program BioAutoGraph (written by J.C. Long and M. Ross), and discrete size categories were assigned to the PCR products.

Deletion of the *DYF155S2* locus (Jobling et al. 1996) was detected by PCR. Primer sequences and PCR conditions were as described by Jobling and Tyler-Smith (1995) and Jobling et al. (1996). A PCR-based assay was also used to type a restriction site polymorphism at the *DYZ3* alphoid satellite DNA locus (Tyler-Smith and Brown 1987; Tyler-Smith et al. 1993; Santos et al. 1995). Forty nanograms of genomic DNA was amplified with 200 μM dNTPs, 10 mM Tris-HCl (pH = 8.3), 50 mM KCl, 1.0–2.0 mM $MgCl_2$, 0.6 U of AmpliTaq polymerase, and 0.33 μM of the forward and reverse primers designated U972 (5′-TCT GAG ACA CTT CTT TGT GGT A-3′) and L1214 (5′-CGC TCA AAA TAT CCA CTT TCA C-3′). The PCR conditions were as follows: 94°C for 3 min, and then 30 cycles at 94°C for 30 s, 65°C for 30 s, and 72°C for 1 min. After amplification, the PCR products were visualized on a 3.5% agarose gel. The presence or absence of the restriction site polymorphism was detected by use of 5 μl of PCR mixture, 20 U of the restriction enzyme *Hin*dIII (New England Biolabs), and the recommended buffer incubated overnight at 37°C.

Phylogenetic and Statistical Analyses

All nine loci were used to construct Y chromosome haplotypes. The phylogeny of Y chromosome haploty-

pes was inferred by maximum parsimony, by means of the program Phylogenetic Analysis Using Parsimony (PAUP 3.1.1; Swofford 1993). Y chromosome microsatellite alleles were defined as ordered characters, incorporating the single stepwise mutation model, after the mutation/drift model of Kimura and Ohta (1978), which has been proposed as a model for microsatellite mutations (Shriver et al. 1993; Valdes et al. 1993; DiRienzo et al. 1994). Confidence in the phylogeny was determined by bootstrapping (10,000) and repeated cladogram constructions. An unrooted haplotype phylogeny with branch lengths proportional to mutational steps was constructed. Major haplogroups were supported by 100% of bootstraps.

Haplotype diversity, the probability that two individuals chosen at random from a population have different Y chromosome haplotypes, was estimated, along with its standard errors, by use of equations (8.5) and (8.13) of Nei (1987). Differences among populations were assessed by use of the hierarchical analysis of molecular haplotype variance calculated with the WINAMOVA computer package (Excoffier et al. 1992; Michalakis and Excoffier 1996). The distance matrix used in the analysis of molecular variance (AMOVA) contained the sum of squared differences in the repeat numbers for each microsatellite locus. Variance components were estimated for three hierarchical levels: Φ_{CT}, groups of Finnish provinces (regions) relative to the total population; Φ_{SC}, provinces relative to groups of provinces (regions); and Φ_{ST}, provinces relative to the total population. Significance levels were obtained by comparison of observed values to the empirical null distribution from 1,000 randomizations.

Results

Phylogenetic Analysis

Table 1 shows the 77 distinct Y chromosome haplotype configurations that were observed in the Finnish population. Thirty-four of the Y haplotypes were observed in single copies. Only two haplotypes (49 and 69) were found in high frequency. Haplotype 49 was observed in 81 males (29%) and haplotype 69 in 30 males (11%) (table 1).

The unrooted haplotype phylogeny relating all 77 Y haplotypes is depicted in figure 1. Three distinct haplogroups were observed. Haplogroup A is the largest (55%), and all members of this group of haplotypes share the *DYF155S2* deletion polymorphism associated with several Asian populations (Jobling et al. 1996). Haplogroups B and C do not possess the deletion, but are defined by the presence or absence of the *Hin*dIII restriction site at the *DYZ3* locus. Haplogroup B, the second largest group of haplotypes (29%), possesses the restriction site. Haplogroup B is phylogenetically complex and is composed of four subgroups. All four subgroups are found in other European populations (unpublished data). However, subgroup B1 appears to have expanded within Finland, whereas subgroups B2, B3, and B4 consist of divergent haplotypes (fig. 1). Y chromosomes from haplogroup C do not possess the *Hin*dIII restriction site. Similar to the three subgroups, B2, B3, and B4, haplogroup C consists of three subgroups of rare divergent haplotypes that most likely represent recent European immigrants into Finland.

Phylogenetic analysis provides evidence for the antiquity of common haplotypes A/49 and B/69. The unrooted haplotype phylogeny in figure 1 depicts both groups as star-shaped with haplotypes coalescing to the most common Y haplotype in the group. A majority of the single-copy haplotypes found in the Finnish population are within four mutational steps from either haplotype A/49 or haplotype B/69. Others are rare divergent haplotypes. This is strong evidence for a recent population expansion from the two common Y haplotypes. Also, $\geqslant$10 microsatellite mutational steps, plus the deletion at the *DYF155S2* locus, separate the two common haplotypes. Thus, it is very unlikely that haplogroups A and B originate from the same source population.

The unrooted haplotype phylogeny reveals that almost all of the haplotypes in haplogroups A and B are linked to other haplotypes by just one single-step mutation. This was especially true for haplogroup A, where all but two of the intermediate haplotypes were observed. These hypothetical haplotypes are represented by slashes (fig. 1). However, haplogroup B contained three subgroups, B2, B3, and B4, which diverged from haplotype B by at least three mutational steps. Most importantly, many of the intermediate haplotypes for haplogroups B and C were not observed in the Finnish population. These observations suggest that subgroups B2, B3, and B4 do not coalesce to the common haplotype B within the Finnish population, but to a copy of haplotype B in another population. Similarly, subgroups C1, C2, and C3 should coalesce to a common haplotype outside of Finland. In fact, each of the intermediate haplotypes for haplogroups B and C are observed in Swedish and Euroamerican males (unpublished data). Thus, either haplotypes in group B were possessed by a diverse group of founders and subgroup B1 was the most successful, or subgroups B2, B3, and B4 were introduced by recent European immigrants. Either way, subgroups B2, B3, and B4 are unlikely to be part of the Finnish expansion from haplotype B/69.

Since haplogroups A and B1 coalesce to a common haplotype, we calculated a rough estimate of the coalescence times of haplotypes within the two groups. We assume that the coalescence times for the two haplogroups co-occur with their entry into Finland. If this is

Table 1

Finnish Y Chromosome Haplotypes

Group	Haplotype	*DYS388*	*DYS389a*	*DYS390*	*DYS391*	*DYS392*	*DYS393*	*DYS394*	*DYZ3*	*DYF155S2*	No. Observed
A	49	129	255	214	287	257	128	247	+	+	81
A	25	129	251	214	287	257	128	247	+	+	6
A	26	129	251	214	287	257	132	247	+	+	5
A	35	129	251	218	287	257	128	259	+	+	2
A	36	129	251	218	287	257	128	247	+	+	1
A	44	129	255	214	283	257	128	259	+	+	1
A	50	129	255	214	287	257	132	247	+	+	4
A	51	129	255	214	287	260	128	247	+	+	1
A	52	129	255	214	287	257	128	251	+	+	2
A	53	129	255	214	287	257	128	255	+	+	1
A	54	129	255	214	287	257	128	259	+	+	2
A	55	129	255	214	287	257	128	263	+	+	2
A	56	129	255	214	291	257	128	247	+	+	2
A	57	129	255	214	291	260	128	247	+	+	1
A	58	129	255	218	287	257	128	247	+	+	4
A	9	129	251	210	283	257	124	247	+	+	2
A	10	129	251	210	283	257	128	247	+	+	5
A	11	129	251	210	283	257	128	251	+	+	2
A	20	129	251	214	283	257	128	247	+	+	1
A	37	129	255	210	283	257	128	247	+	+	4
A	43	129	255	214	283	257	128	251	+	+	1
A	45	129	255	214	283	257	128	247	+	+	8
A	59	129	259	210	283	257	128	247	+	+	1
A	60	129	259	214	283	257	132	247	+	+	1
A	12	129	251	210	287	257	120	247	+	+	2
A	13	129	251	210	287	257	124	247	+	+	1
A	14	129	251	210	287	257	128	247	+	+	11
A	16	129	251	210	287	257	128	251	+	+	1
A	24	129	251	214	287	257	124	247	+	+	1
A	39	129	255	210	287	257	124	247	+	+	2
A	40	129	255	210	287	257	128	247	+	+	7
A	41	129	255	210	287	257	132	247	+	+	1
A	48	129	255	214	287	257	124	247	+	+	1
A	62	132	251	210	287	257	124	247	+	+	1
B1	69	135	247	210	283	248	124	247	+	...	30
B1	61	132	247	210	283	248	124	247	+	...	1
B1	65	135	247	206	283	248	124	247	+	...	2
B1	66	135	247	206	283	248	124	251	+	...	1
B1	67	135	247	206	283	251	124	251	+	...	1
B1	68	135	247	210	283	248	124	243	+	...	2
B1	70	135	247	210	283	248	124	251	+	...	4
B1	71	135	247	210	283	248	124	259	+	...	2
B1	72	135	247	210	283	251	124	247	+	...	3
B2	73	141	251	210	283	248	120	247	+	...	3
B2	74	141	255	214	283	257	132	247	+	...	2
B2	75	141	255	214	283	257	128	247	+	...	1
B2	76	144	247	210	283	248	124	247	+	...	1
B2	77	144	251	210	283	248	124	251	+	...	1
B2	78	144	255	214	283	257	128	247	+	...	2
B3	21	129	251	214	287	248	124	259	+	...	1
B3	28	129	251	218	283	248	124	263	+	...	1
B3	29	129	251	218	283	248	124	255	+	...	1
B3	30	129	251	218	283	248	128	251	+	...	2
B3	63	132	251	214	283	248	124	255	+	...	1
B3	64	132	251	214	287	248	124	251	+	...	2
B4	5	129	247	210	283	248	124	247	+	...	7
B4	6	129	247	210	283	248	124	251	+	...	1
B4	7	129	251	210	283	248	124	251	+	...	2
B4	18	129	251	214	283	248	124	243	+	...	1
B4	38	129	255	210	287	248	128	247	+	...	1
C1	3	126	251	214	287	254	124	247	...	...	1

(*continued*)

Table 1 (continued)

Group	Haplotype	*DYS388*	*DYS389a*	*DYS390*	*DYS391*	*DYS392*	*DYS393*	*DYS394*	*DYZ3*	*DYF155S2*	No. Observed
C1	4	129	243	214	283	254	124	247	...	...	2
C1	8	129	251	210	283	254	120	251	...	...	2
C1	19	129	251	214	283	254	124	247	...	...	3
C1	22	129	251	214	287	254	124	247	...	...	4
C1	23	129	251	214	287	254	128	247	...	...	1
C1	34	129	251	218	287	254	124	247	...	...	1
C1	42	129	255	214	283	257	124	247	...	...	2
C1	47	129	255	214	287	254	124	247	...	...	3
C2	1	126	247	210	283	248	124	247	...	...	1
C2	2	126	251	206	283	248	124	243	...	...	2
C2	17	129	251	214	283	248	124	247	...	...	1
C3	27	129	251	218	283	248	124	251	...	...	3
C3	31	129	251	218	287	248	124	259	...	...	2
C3	32	129	251	218	287	248	124	255	...	...	3
C3	33	129	251	218	287	248	124	251	...	...	4
C3	46	129	255	214	287	248	128	255	...	...	1

NOTE.—Allele sizes (bp) are shown for the seven microsatellite loci. A plus sign (+) at *DYZ3* and *DYF155S2* denotes the presence of the *Hin*dIII or deletion polymorphisms at the respective locus.

so, then the expected variance in microsatellite repeat length (S) can be used to estimate T, the expected value of the average coalescence time for microsatellite alleles or time since rapid population growth. The theoretical relationship $S = 2\mu T\sigma^2$ was applied using a mutation rate (μ) of 10^{-3} and a mutational variance (σ^2) of 1 for the one-step mutation model (DiRienzo et al. 1994; Slatkin 1995). By use of the observed variance in S of 0.39 for haplogroup A and 0.13 for haplogroup B1, and a generation time of 25 years, the approximate coalescence times for haplogroups A and B were found to be 4,875 and 1,625 years ago, respectively. Coalescence times were also calculated by use of the recently estimated mutation rate of 2.1×10^3 (95% confidence interval [CI] 0.6×10^{-3} and 4.9×10^{-3}) for Y chromosome–specific microsatellites (Heyer et al. 1997). The estimates were 2,321 years ago (95% CI 8,125–995 years ago) for haplogroup A and 774 years ago (95% CI 2,708–332 years ago) for haplogroup B. Nevertheless, the 95% CIs for the coalescence times do not conflict with the archeological data for the arrival of two distinct groups of settlers.

Y Haplotype Diversity and Distribution

Haplotype diversity was estimated at 0.899 ± 0.007 for the entire population. This value is quite striking, especially since studies using fewer microsatellite loci have observed significantly higher Y chromosome haplotype diversity in other European populations (Cooper et al. 1996; Roewer et al. 1996). Table 2 details Y haplotype diversity for each of the provinces. These values range from 0.971 ± 0.004 in the urban Turku (2) province to 0.748 ± 0.031 in the eastern province of Northern Karelia (9).

The frequencies of haplogroups A and B for each province are also presented in table 2. Since haplogroup A is the most common throughout Finland, it can be inferred that entry of this haplogroup into the Finnish population, and subsequent expansion, predate haplogroup B. We have been able to determine that haplogroup B is not exclusive to Finland but is also found in Swedish ($n = 45$) and Euroamerican ($n = 32$) males at frequencies of ~16% and ~5%, respectively. In sharp contrast to haplogroup B, haplogroup A is a divergent cluster of related haplotypes not found in Swedish and Euroamerican males.

Figure 2 depicts the geographic distribution of haplotypes A/49 and B/69. Interestingly, haplotype A/49 was found in highest frequency in the eastern provinces of Northern Karelia, Kuopio, and Kymi. The most striking of these provinces is Northern Karelia, where 50% of males possess haplotype A/49, whereas only 2% possess haplotype B/69. The opposite pattern is observed in the southwest. Table 2 shows that haplotype A/49 was observed in Turku at 6% and Helsinki at 29%, yet its descendant haplotypes (haplogroup A) were observed at a much higher frequency (58% and 42%, respectively). This pattern most probably reflects recent population movement into urban areas during the period of industrialization after World War II (Norio 1981).

Finnish Population Substructure

The apportionment of Y chromosome haplotype diversity in Finnish provinces was performed by use of AMOVA. Provinces were grouped together into regions that distinguish the early and late settlements of Finland (de la Chapelle 1993). The early settlements consisted of the Helsinki (1), Turku (2), Hame (3), and Kymi (4)

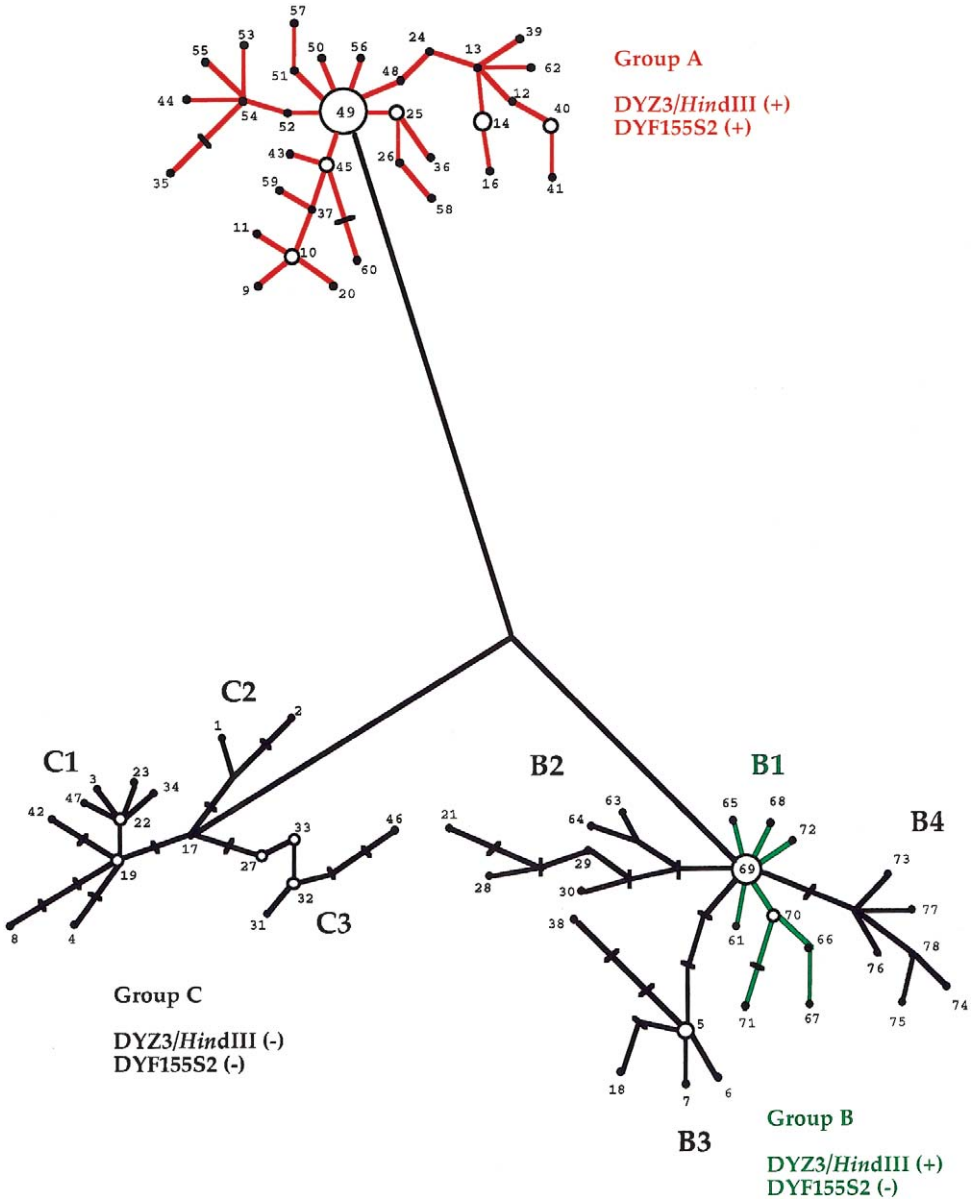

Figure 1 Unrooted Y haplotype phylogeny relating 77 Y chromosome haplotypes observed within the Finnish population. Haplogroups A and B1 are depicted in red and green, respectively. Branch lengths are proportional to the minimum number of mutational steps separating haplotypes. Numbers represent observed haplotypes. Circles are proportional to haplotype frequencies.

provinces; whereas the late settlements included Vaasa (6), Kuopio (8), Northern Karelia (9), Oulu (10), and Lappi (11) (fig. 2). Table 3 shows that 10% of the total genetic variance is attributable to differences between the Finnish provinces ($P < .001$). A significant amount (~4%, $P < .02$) of the total variance was due to differences between the early and late settlements of Finland. The Φ_{SC} estimate of 0.065 reveals that an appreciable amount of Y chromosome divergence was also observed within the two regions.

Discussion

Convincing genetic evidence for the dual origins of Finns was revealed by an analysis of Y chromosome compound haplotypes from nine Finnish provinces. Two distinct male lineages contributed to the present Finnish population. Haplotypes were constructed from nine Y-specific loci, including a deletion polymorphism (*DYF155S2*) found in populations of Asian ancestry. We note that the deletion polymorphism was found by Jo-

Table 2

Summary of Y Chromosome Haplotype Frequencies and Diversity for Finnish Provinces

Province (Code)	*n*	*k*	Haplotype A	Haplotype B	Group A	Group B	Haplotype Diversity
Helsinki (1)	34	16	.29	.24	.42	.41	.866 ± .020
Turku (2)	51	32	.06	.12	.58	.27	.971 ± .004
Hame (3)	20	13	.20	.15	.50	.25	.942 ± .013
Kymi (4)	20	7	.40	.25	.45	.55	.784 ± .030
Vaasa (6)	20	14	.15	.15	.50	.40	.958 ± .009
Kuopio (8)	37	20	.38	.03	.76	.16	.856 ± .027
Northern Karelia (9)	58	22	.50	.02	.83	.15	.748 ± .031
Oulu (10)	20	12	.25	.00	.50	.15	.921 ± .018
Lappi (11)	20	11	.25	.15	.40	.25	.918 ± .018
Total (pooled)	280	77	.29	.11	.55	.29	.899 ± .007

NOTE.—n = number of chromosomes, and k = observed number of haplotypes.

bling et al. (1996) in one Norwegian and two Greeks. Since the polymorphism is rare in these two populations, its presence is most likely due to recent male migrants with Asian ancestry. Similar results were observed by Zerjal et al. (1997) in relation to a T→C transition on the Y chromosome. In fact, the C allele chromosomes were found to compose a subset of *DYF155S2* deleted chromosomes (Zerjal et al. 1997).

In contrast to other Scandinavian populations, the most common and widely distributed haplotype (A/49) in Finns contained the *DYF155S2* deletion. Haplotype A/49 is restricted to Finland and has not been found in other Scandinavian populations (unpublished data). The other common Y haplotype (B/69) lacks the deletion polymorphism and is observed in other European populations. From the geographical distribution and molecular divergence of haplotypes A/49 and B/69, we conclude that they represent two major founding Y chromosome lineages in Finland. Haplotype A/49 is most common in the northeast, near the Lake Ladoga region, and haplotype B/69 is common in the south. A radiation or movement of males from these two regions toward the interior of Finland may have occurred, since both haplotypes A/49 and B/69 were observed at approximately equal frequencies in the Hame (3), Vaasa (6), and Kymi (4) provinces. An interesting pattern was observed in the Lappi (11) and Oulu (10) provinces. Haplotype B/69 was not detected in the Oulu province; however, just north, in the Lappi province, it was observed at a frequency of 25%. Haplotype B/69 may have entered the Lappi province from Sweden, since the western border of Lappi is shared with Sweden, whereas the Gulf of Bothnia is west of the Oulu province.

The geographic distribution of haplotypes depicts separate and distinct routes into Finland for the two groups of settlers. Haplotype A/49 may have entered Finland from the Lake Ladoga region, whereas settlers possessing haplotype B/69 traveled across the Gulf of Finland and the Swedish border. The multiple routes of entry for haplotype B/69, in addition to the complex phylogeny of haplogroup B, suggests that the expansion of haplogroup B in Europe predates its entry into Finland. We cannot rule out the possibility that haplotype B/69 is also common in Asia. To date however, we have not observed the haplotype in Asians (Taiwanese, $n = 20$, and Koreans, $n = 20$).

Although it is difficult to determine the level of variation that existed within the two founding lineages at the time of settlement, the two groups were unlikely to be homogeneous for their Y chromosomes. Estimates of the time of expansion for the two haplogroups reveal that haplotype A/49 predates haplotype B/49 in Finland. This is concordant with the high frequency and wide geographic distribution of haplotype A/49. The estimated expansion time for haplogroup A predates that for haplogroup B by almost 2,000 years. Although a mutation rate has recently been estimated for Y chromosome microsatellites (Heyer et al. 1997), the mutational variance has not been formally established. If we include this uncertainty along with others, such as the size, level of variation, and rate of growth of the founding populations, the 95% CIs for the time estimates are expanded further. Nevertheless, the expansion times are informative and provide clear support for separate foundings of the population. This is important because the founding of Finland ~2,000 years ago is the defining characteristic of the single-origin model. Our results argue against the single-origin model. Specifically, we find evidence that the initial group of settlers provided a substantial contribution (55%) to the present Finnish Y chromosome gene pool. These settlers were of Asian ancestry and were followed by a second, genetically distinct wave of settlers. This second group, possessing haplogroup B, may have arrived in Finland with the wave of agriculturists who shaped much of the genetic landscape of Europe (Sajantila and Paabo 1995). These estimates are consistent with archeological data that suggest that the first settlers were Uralic speakers who arrived ~4,000

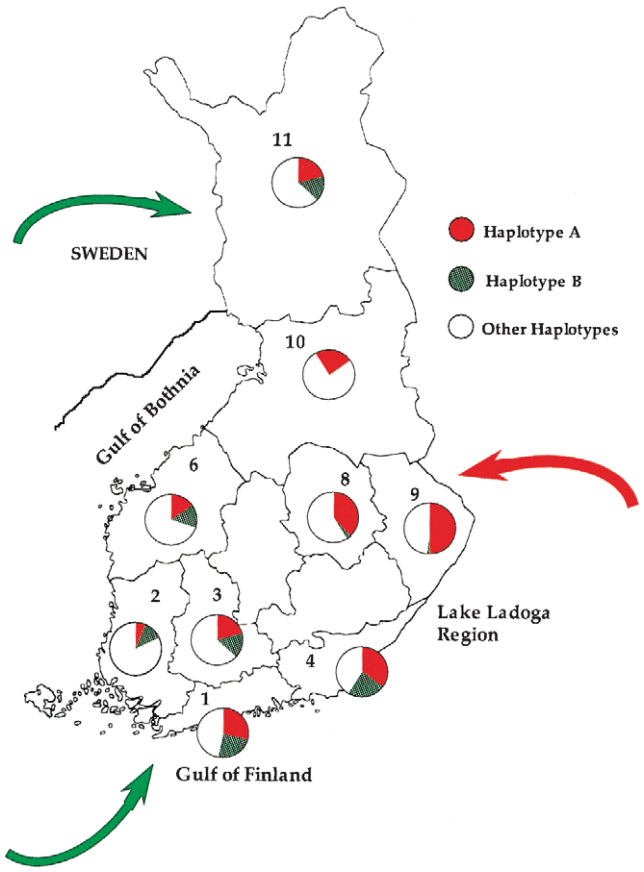

Figure 2 Geographical distribution of common haplotypes A and B for nine provinces within Finland. 1 = Helsinki, 2 = Turku, 3 = Hame, 4 = Kymi, 6 = Vaasa, 8 = Kuopio, 9 = Northern Karelia, 10 = Oulu, and 11 = Lappi. Arrows represent probable points of entry of settlers into Finland.

years ago (Fodor and Czeizel 1991) and that a later group settled along the southern shores ~2,000 years ago (Luho 1976).

Partition of Y chromosome variance revealed that 10% of the total variance was found between provinces. Interestingly, ~4% of the total variance is attributed to differences between what have been termed "old Finland" and "new Finland" (de la Chapelle 1993). These values are quite high for a population traditionally viewed as a homogeneous genetic isolate. In fact, the significant genetic differences between provinces are due not only to the mixed or dual origins of Finns but also to the limited number of settlers and isolation by distance.

We note that, since genetic drift has been a strong force operating within Finland (Nevanlinna 1972; Norio 1981; de la Chapelle 1993; Peltonen et al. 1995), it may also be responsible for elevated Y haplotype frequencies in certain regions. However, if the geographical pattern of Y haplotypes is due to drift only, the enrichment would be expected to be random, and no pattern would be expected, since different Y haplotypes would be enriched in different regions. Nevanlinna (1972) has shown that genetic drift within Finland operated randomly across each subisolate resulting in the loss and enrichment of rare genes.

Table 3

Genetic Differentiation of Finnish Provinces

Type of Comparison	Percent Variance	Φ statistic	*P* value
Among regions (old vs. new Finland)	3.76	Φ_{CT} = .038	.05
Among provinces/ within regions	6.24	Φ_{SC} = .065	.001
Within provinces	90.01	Φ_{ST} = .100	<.001

In summary, our interpretation of the Y chromosome data is that two separate founder populations provided a substantial contribution to the Finnish gene pool. A strong clinal distribution of Y haplotype frequencies was observed, indicative of a migration of males from the Lake Ladoga region into the interior of Finland. This pattern contradicts predictions of the single-origins hypothesis of a single homogeneous group of founders arriving in the south and slowly dispersing into the north (de la Chapelle 1993; Lahermo et al. 1996). Concordant with the archeological evidence for the dual-origins hypothesis, significant differences between eastern and western Finland exist for two common yet polyphyletic Y chromosome haplotypes.

Acknowledgments

We would like to thank Dr. Erkki Vartiainen, for DNA samples, and Drs. Jaakko Lappalainen and S.O.Y. Keita, for helpful discussions. We also appreciate the comments from two reviewers.

References

Cooper G, Amos W, Hoffman D, Rubinsztein D (1996) Network analysis of human Y microsatellite haplotypes. Hum Mol Genet 5:1759–1766

de la Chapelle A (1993) Disease gene mapping in isolated human populations: the example of Finland. J Med Genet 30:857–865

Di Rienzo A, Peterson A, Garza J, Valdes A, Slatkin M, Freimer N (1994) Mutational processes of simple-sequence repeat loci in human populations. Proc Natl Acad Sci USA 91: 3166–3170

Eriksson A (1973) Genetic polymorphism in Finno-Ugrian populations: Finns, Lapps, and Maris. Israel J Med Sci 9: 1156–1170

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA hap-

lotypes: application of human mitochondrial DNA restriction data. Genetics 131:479–491
Fodor I, Czeizel A (1991) The origins of the Hungarian population. In: Czeizel A, Benkmann H, Goedde H (eds) Genetics of the Hungarian population. Springer-Verlag, Berlin
Hajdu P (1975) Finno-Ugrian languages and peoples. Andre Deutsch Ltd, London
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
Jobling M, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, et al (1996) Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum Mol Genet 5: 1767–1775
Jobling M, Tyler-Smith C (1995) Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449-456
Kimura M, Ohta T (1978) Stepwise mutation model and the distribution of allelic frequencies in a finite population. Proc Natl Acad Sci USA 75:2868–2872
Kittles RA, Urbanek M, Goldman D, Long JC (1996) Microsatellite polymorphisms reveal average heterozygosity for Finns comparable with other European populations. Am J Phys Anthropol Suppl 22:139
Lahermo P, Sajantila A, Sistonen P, Lukka M, Aula P, Peltonen L, Savontaus M (1996) The genetic relationship between the Finns and the Finnish Saami (Lapps): analysis of nuclear DNA and mtDNA. Am J Hum Genet 58:1309–1322
Luho V (1976) The population and prehistory of Finland. In: Hajdu P (ed) Ancient cultures of the Uralian peoples. Corvina Press, Budapest, pp 115–132
Meinander CF (1973) The problem of the Finno-Ugrian peoples' origin on the base of archaeological data. In: Studies in the anthropology of the Finno-Ugrian peoples, Stencil no. 7. Archaeology Institute, University of Helsinki, Helsinki
Michalakis Y, Excoffier L (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. Genetics 142: 1061–1064
Nevanlinna HR (1972) The Finnish population structure: a genetic and genealogical study. Hereditas 71:195–236
——— (1973) Further evidence of the inhabitation in a marginal population. Hereditas 74:127–132
Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York
Norio R (1981) Diseases in Finland and Scandinavia. In: Rothschild H (ed) Biocultural aspects of disease. Academic Press, New York, pp 359–415
Norio R, Nevanlinna H, Perheentupa J (1973) Hereditary diseases in Finland: rare flora in rare soil. Ann Clin Res 5: 109–141
Peltonen L, Pekkarinen P, Aaltonen J (1995) Messages from an isolate: lessons from the Finnish gene pool. Biol Chem Hoppe-Seyler 376:697–704
Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome–specific microsatellites in two closely related human populations. Hum Mol Genet 5: 1029–1033
Ruhlen M (1987) A guide to the world's languages. Vol 1. Stanford University Press, Stanford
Sajantila A, Paabo S (1995) Language replacement in Scandinavia. Nat Genet 11:359–360
Sajantila A, Salem A, Savolainen P, Bauer K, Gierig C, Paabo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. Proc Natl Acad Sci USA 93:12035–12039
Santos F, Pena S, Tyler-Smith C (1995) PCR haplotypes for the human Y chromosome based on alphoid satellite DNA variants and heteroduplex analysis. Gene 165:191–198
Shriver M, Jin L, Boerwinkle E, Deka R, Ferrell R, Chakraborty R (1995) A novel measure of genetic distance for highly polymorphic tandem repeat loci. Mol Biol Evol 12: 914–920
Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139:457–462
Swofford D (1993) PAUP, version 3.1.1. Smithsonian Institution, Washington, DC
Tyler-Smith C, Brown W (1987) Structure of the major block of alphoid satellite DNA on the human Y chromosome. J Mol Biol 195:457–470
Tyler-Smith C, Oakley RT, Larin Z, Fisher R, Crocker M, Affara N, Ferguson-Smith M, et al (1993) Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. Nat Genet 5:368–375
Valdes A, Slatkin M, Freimer N (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749
Vartiainen E, Puska P, Jousilahti P, Korhonen H, Tuomilehto J, Nissinen A (1994) Twenty-year trends in coronary risk factors in North Karelia and in other areas of Finland. Int J Epidemiol 23:495–504
Vilkuna K (1976) The formation of the Finnish peasant way of life. In: Hajdu P (ed) Ancient cultures of the Uralian peoples. Corvina Press, Budapest, pp 183–190
Virkkunen M, Kallio E, Rawlings R, Tokola R, Poland RE, Guidotti A, Nemeroff C, et al (1994) Personality profiles and state aggressiveness in Finnish, alcoholic, violent offenders, fire setters, and healthy volunteers. Arch Gen Psychiatry 51:28–33
Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos F, Schiefenhovel W, et al (1997) Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183