AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 108:381–399 (1999)

Autosomal, Mitochondrial, and Y Chromosome DNA Variation in Finland: Evidence for a Male-Specific Bottleneck

RICK A. KITTLES,[1,2]* ANDREW W. BERGEN,[3] MARGRIT URBANEK,[1] MATTI VIRKKUNEN,[5] MARKKU LINNOILA,[4] DAVID GOLDMAN,[3] AND JEFFREY C. LONG[1]
[1]*Section on Population Genetics and Linkage, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892*
[2]*Department of Biological Sciences, George Washington University, Washington, DC 20052*
[3]*Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892*
[4]*Laboratory of Clinical Studies, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, Maryland 20892*
[5]*Department of Psychiatry, University of Helsinki, Helsinki, Finland 00180*

KEY WORDS Y-chromosome compound haplotypes; Y-haplotype phylogeny; microsatellite DNA; mtDNA control region; population bottleneck

ABSTRACT The high prevalence of rare genetic diseases in Finland has been attributed to a founder effect some 2,000 years ago. However, this hypothesis has not been supported from mtDNA sequence and autosomal microsatellite data which indicate high levels of gene diversity. Here we have identified genetic evidence for a population bottleneck by examining variable microsatellite loci on the nonrecombining portion of Y chromosomes from Finland and four populations from Europe and the Americas. Sequence data from segment I of the control region (HVS-1) of mtDNA (360 bases) and 20 autosomal dinucleotide repeat markers were also analyzed. Partitions of genetic variance within and between populations revealed significant levels of Y-chromosome differentiation between populations. Phylogenetic and diversity analyses revealed divergent Finnish Y-haplotype clades and significantly lower Y-haplotype diversity among Finns as compared to other populations. Surprisingly, Finnish Y-haplotype diversity was even lower than the Native American populations. These results provide support for the Finnish bottleneck hypothesis. Evidence for two separate founding Finnish Y-chromosome lineages was also observed from the Y-chromosome phylogeny. A limited number of closely related founding males may have contributed to the low number of paternal lineages in the Finnish population. In contrast, high levels of genetic diversity for mtDNA and autosomal STRs may be the result of sex-biased gene flow and recent immigration to urban areas from established internal isolates within Finland. Am J Phys Anthropol 108:381–399, 1999.

Geneticists have considered the Finnish population to be an isolate founded some 2,000 years ago (Nevanlinna, 1972). Within the last 300 years, the population has expanded from an estimated 250,000 to more than 5 million. Throughout this period of rapid expansion, Finland has been relatively isolated from neighboring countries for both geographic and sociocultural reasons. Finns

*Correspondence to: Rick Kittles, Howard University Cancer Center, 2041 Georgia Ave., Rm. 507, Washington, DC 20060. E-mail: rkittles@howard.edu

Received 14 November 1997; accepted 10 December 1998.

do not speak an Indo-European language as do most other European populations. Their language, Finnish, is a branch of the Uralic language family (Ruhlen, 1987). Within the Finnic branch are closely related languages spoken by Finns, Estoninans, and Saami (Lapps).

Two models have been proposed for the origin of the Finns. The single-origin model purports a founding of Finland some 2,000 years ago by a small number of settlers (Nevanlinna, 1972; de la Chapelle, 1993), followed by relative isolation. 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 of settlers from the south via the Gulf of Finland (Norio, 1981; Meinander, 1973; Eriksson, 1973). Similar to the opposing model, the dual-origin model asserts that the Finnish population has remained isolated for some 2,000 years.

The purported consequences of the founding of the Finnish population and subsequent isolation are a high prevalence of recessive diseases (>30) which are otherwise rare in other parts of the world (de la Chapelle, 1993), reduced heterozygosity (gene diversity) of the Finnish population, and genetic differentiation of Finns from other European populations. Reduced heterozygosity is the characteristic signature of population bottlenecks (Nei et al., 1975). However, reduced heterozygosity in Finns has never been demonstrated. Although some comparative studies of classical blood antigen and protein loci suggest that Finns are distinct outliers from other Europeans (Erikkson, 1973; Sokal et al., 1988, 1989; Guglielmino et al., 1990; Cavalli-Sforza and Piazza, 1993; Nei and Roychoudhury, 1993; Sanchaz-Mazas et al., 1994), recent studies using nuclear DNA marker (mini- and microsatellites) loci have not confirmed these findings (Sajantila et al., 1992; Lahermo et al., 1996; Kittles et al., 1996). Consequently, the nuclear DNA marker data do not support the bottleneck hypothesis.

Although the population bottleneck hypothesis has been used in the past to explain the Finnish population structure, the hypothesis has received little scrutiny. In fact, the bottleneck hypothesis is contradicted by the uneven geographic distribution of genetic recessive diseases in Finland. This distribution of genetic diseases is better explained by isolation by distance or "internal isolates" each with its own signature of enriched recessive alleles and perhaps an overall increase in recessive diseases. It is important to reconstruct the demographic history of the Finnish population since the population has been successfully used in gene-mapping studies (Hastbacka et al., 1992; Hoglund et al., 1995; Varilo et al., 1996; see also Peltonen et al., 1995). Much of this success has been attributed to Finland's unique population history and excellent church and family records (de la Chapelle, 1993; Hastbacka et al., 1992). Thus, there is a need to evaluate how the Finnish population structure has facilitated gene-mapping efforts. For instance, in the search for founder disease mutations for complex disorders within Finland, it may be important that families with the disease in question originate from the same geographic area. This may decrease the level of genetic heterogeneity in the disease sample. Here, for the first time, data from different polymorphic systems consisting of microsatellite loci from autosomal and Y chromosomes and mtDNA control region sequences are used together to evaluate patterns of genetic variation and test the hypothesis of a bottleneck event during the founding of the Finnish population. These markers have higher heterozygosities than classical markers and thus are more informative for assessing population structure.

Autosomal microsatellite loci have been utilized for population studies (Bowcock et al., 1994; DiRienzo et al., 1994; Deka et al., 1995; Goldstein et al., 1995; Shriver et al., 1995; Jorde et al., 1995; Nei and Takezaki, 1996), mainly due to an ease in typing using high throughput polymerase chain reaction (PCR) technology. A vast majority of these loci have $>70\%$ heterozygosity levels in Europeans (Deka et al., 1995) and thus are very powerful for genetic linkage and microevolutionary studies.

Polymorphic genetic elements with sex-specific transmission have also proven to be extremely useful and informative in popula-

tion studies. Segment I (positions 16024–16383) of the mtDNA control region has been extensively used by investigators for population studies (see Bertranpetit et al., 1995; Mountain et al., 1995; DiRienzo and Wilson, 1991; Ward et al., 1991; Sajantila et al., 1995; Graven et al., 1995), and the utility of the Y chromosome has been increasing since more paternally inherited single nucleotide and microsatellite polymorphisms have been characterized on the nonrecombining portion of the chromosome (Hammer et al., 1997; Jobling and Tyler-Smith, 1995; Santos et al., 1996; Underhill et al., 1997). Recent analyses of Y-chromosome microsatellites have revealed high levels of variation in closely related human populations (Deka et al., 1996; Roewer et al., 1996). The lack of recombination for Y-specific microsatellites and mtDNA allows for easy construction of definitive haplotypes which provide information on paternal and maternal lineages. This study assesses the levels of gene diversity and differentiation for autosomal DNA loci and paternal and maternal lineages from three populations of European origin and two Native American populations. The pattern of genetic variation in Finns was compared to that of the other two European and two Native American populations. These four were ideal comparative populations since the European populations had not undergone a recent population bottleneck, in contrast to the Native American populations. For the Finns, a major reduction in Y-chromosome variation was observed, in support of a population bottleneck.

MATERIALS AND METHODS

Sample collection and processing

Eighty-nine unrelated males from diverse geographic regions of Finland were ascertained as volunteers from newspaper advertisements for a psychiatric genetic study in Helsinki, Finland. All males were Finnish speakers. In addition, unrelated males from Lund, Sweden (n = 32), European Americans from Bethesda, Maryland (n = 45), and two Native American populations, Pima (n = 55) and Cheyenne (n = 31) were included. Native Americans were identified based on their eligibility for tribal enrollment. Individuals of mixed tribal or ethnic ancestry were not excluded. Both the Pima and Cheyenne are members of the Amerind language family according to Greenberg's new classification (1987). All samples were collected with informed consent. Samples from each of the five populations were typed for 20 unlinked autosomal (dinucleotide repeat) and seven male-specific Y-chromosome (5 tetra- and 2 trinucleotide repeat) microsatellites and a restriction site at the Y-chromosome alphoid satellite DNA *DYZ3* locus and sequenced for HVS-1 of the mtDNA control region.

All microsatellite loci were typed following PCR amplification with fluorescently labeled upper primers. For each PCR reaction, 50 ng of DNA was added to 200 μM of dNTPs, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.0–2.0 mM $MgCl_2$, 0.6 units of AmpliTaq polymerase (Perkin Elmer, Foster City, CA), and 0.33 uM of primers. PCR cycling conditions were 93°C for 3 min, 10 cycles at 94°C for 15 sec, 55°C for 15 sec, and 72°C for 30 sec, followed by 20 cycles at 89°C for 15 sec, 55°C for 15 sec, and 72°C for 30 sec. The final extension cycle was at 72°C for 10 min. The PCR products were then pooled in the presence of a size standard and electrophoresed using an Applied Biosystems, Inc. (ABI) (Foster City, CA), 373A DNA sequencer.

The autosomal microsatellite loci we evaluated are unlinked and located on three chromosomes: *D1S196*, *D1S197*, *D1S206*, *D1S213*, *D1S220*, *D1S228*, *D1S234*, *D1S235*, *D1S249*, *D1S255*, *D1S484*, *D2S117*, *D2S134*, *D2S157*, *D5S210*, *D5S406*, *D5S407*, *D5S419*, *D5S422*, *D5S433*. Seven Y-specific microsatellite loci were amplified: *DYS388*, *DYS389*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, *DYS394*. The primers for *DYS389* yielded two amplification products, a smaller (240–260 bp) and a larger (370–400 bp) product. Only the smaller product was used in the analysis. Primer sequences for all microsatellite markers used in the study can be found at the Genome Database URL (http://www.gdb.org).

A PCR-based assay was used to type the Y-chromosome alphoid satellite *DYZ3* restriction site (Santos et al., 1995). Forty nanograms of genomic DNA was amplified using 200 μM of dNTPs, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.6

units of AmpliTaq polymerase, and 0.33 µM of the forward and reverse primers designated U972 (5′-TCTGAGACACTTCTTTGTGGTA-3′) and L1214 (5′-CGCTCAAAATATCCACTTTCAC-3′). The PCR conditions were as follows: 94°C for 3 min and then 30 cycles at 94°C for 30 sec, 65°C for 30 sec, and 72°C for 1 min. Following amplification, the PCR products were visualized on a 3.5% agarose gel, and the presence or absence of the restriction site was detected using 20 units of the restriction enzyme *Hin*dIII (New England Biolabs, Beverly, MA).

Hypervariable segment I of the control region of mtDNA was amplified using two oligonucleotide primer sequences from Vigilant et al. (1989): L15996 (5′-CTCCACCATTAGCACCCAAAGC-3′) and H16401 (5′-TGATTTCACGGAGGATGGTG-3′). Amplification was performed using 2 ng of DNA in 150 µM dNTPs, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.0–2.0 mM $MgCl_2$, 0.6 units of AmpliTaq polymerase, and 1 ul of a 5 uM primer mix. The PCR conditions consisted of 95°C for 1 min and 36 cycles of 95°C for 10 sec, 55°C for 30 sec, and 72°C for 30 sec. Both DNA strands were then sequenced using fluorescent-labeled dideoxy terminator cycle sequencing chemistry (ABI) and the ABI 373A DNA sequencer.

Statistical and phylogenetic methods

Genotypes were determined using the GS Analysis and Genotyper programs (ABI). Corrections were made for individual gel shifts using the computer program BioAutoGraph (Long and Ross, 1995). Afterwards, discrete size categories were assigned to the PCR products, and the number of alleles per locus were determined. For the sequence data, the Seq A and AutoAssembler programs (ABI) were used. These programs aligned and overlapped both sequenced strands of DNA, allowing for the visual inspection of ambiguities in the sequence.

Microsatellite allele frequencies were estimated by direct gene counting. Average gene diversity (H) and its standard error were calculated using equations 8.6 and 8.7 of Nei (1987). Y chromosomes were scored as haplotypes using the seven Y-chromosome microsatellite loci and the alphoid repeat polymorphism. Haplotype diversity (h), the probability that two individuals chosen at random from a population have different Y-chromosome or mtDNA haplotypes, was estimated along with its standard error using equations 8.5 and 8.13 of Nei (1987). Nucleotide diversity (π), the measure of gene diversity at the nucleotide position level was estimated for the mtDNA sequence data using equations 10.6 and 10.7 of Nei (1987).

Differences among populations were assessed using two methods, the hierarchical analysis of molecular haplotype variance calculated using the WINAMOVA computer package for Y-chromosome and mtDNA data (Excoffier et al., 1992; Michalakis and Excoffier, 1996) and Slatkin's (1995) measure of population subdivision based on variance of microsatellite repeat length ($\boldsymbol{R}_{ST}$). Estimates using microsatellite data assumed the single stepwise mutation model (Valdes et al. 1993; DiRienzo et al. 1994; Shriver et al. 1993), after the mutation/drift model of Kimura and Ohta (1978). The infinite alleles model (Kimura and Crow, 1964) was assumed for the mtDNA sequence data. For Slatkin's (1995) measure, $\boldsymbol{R}_{ST} = (S - Sw)/S$, where S is twice the estimated variance of microsatellite repeat length for all the populations and Sw is twice the estimated variances of microsatellite repeat length within each population. $\boldsymbol{R}_{ST}$ is the proportion of the total variance in microsatellite repeat length that is due to differences between populations. This measure is analogous to Wright's (1951) $\boldsymbol{F}_{ST}$ and Excoffier et al.'s (1992) $\boldsymbol{\Phi}_{ST}$.

Maximum parsimony (MP) was performed using PAUP 3.1.1 (Swofford, 1993) in order to reconstruct the Y-chromosome haplotype phylogeny. Y-chromosome loci were defined as ordered characters incorporating a single stepwise mutation model (DiRienzo et al., 1994; Valdes et al., 1993). Since mutations were more prevalent among the seven microsatellite loci, equal weights of one were used for the microsatellite loci, while a weight of six was used for the less variable DYZ3 *Hin*dIII restriction site.

Phylogeny reconstruction using mtDNA control region sequences has revealed star topologies for populations which have re-

TABLE 1. Genetic variation at 20 autosomal microsatellite loci in five populations[1]

Population	n	k	H	Sw
Finland	71	9.0	0.779 ± 0.02	44.2 ± 0.52
Sweden	44	9.4	0.786 ± 0.02	44.6 ± 0.48
Euroamerican	32	8.8	0.790 ± 0.01	43.4 ± 0.62
Cheyenne	30	8.6	0.728 ± 0.03	43.0 ± 0.59
Pima	50	7.6	0.710 ± 0.02	42.6 ± 0.52
Overall	227	12.4	0.810 ± 0.02	46.0 ± 0.27

[1] H, average gene diversity; k, average number of alleles; n, sample size; Sw, twice the mean variance in microsatellite repeat length.

TABLE 2. mtDNA diversity in five populations

Population	n	k	h	π
Finland	73	50	0.964 ± 0.007	0.011 ± 0.001
Sweden	28	13	0.839 ± 0.031	0.009 ± 0.001
Euroamerican	44	33	0.965 ± 0.009	0.009 ± 0.001
Cheyenne	39	26	0.973 ± 0.004	0.021 ± 0.002
Pima	40	22	0.932 ± 0.011	0.019 ± 0.002

h, haplotype diversity; k, number of haplotypes; n, sample size; π, nucleotide diversity.

cently grown in size (DiRienzo and Wilson, 1991; Rogers and Jorde, 1995). In theory, a star topology should also be observed for Y-chromosome haplotypes defined by highly variable microsatellites alleles. Coalescence theory suggests that common, geographically dispersed haplotypes should be found at internal nodes of the tree and ancestral to the less frequent, geographically restricted haplotypes (Hudson, 1990; Griffiths and Tavare, 1994; Donnelly, 1996; Templeton, 1993). Thus, we defined major Y-lineage groups as monophyletic clades of haplotypes whose most recent common ancestor (MRCA) was observed at a frequency greater than 4%.

RESULTS

Autosomal microsatellite diversity

Table 1 presents the average number of alleles per autosomal microsatellite locus for the five populations. The average number of alleles for Finns was within the range observed in the other European populations. The Native American populations possessed fewer alleles per locus than the Europeans, possibly due to founder effects during the colonization of the New World and subsequent drift following their divergence as separate populations in North America. Gene diversity for each locus and population is shown in Table 1. The Euro-American sample exhibited the highest diversity (0.790 ± 0.01), while the Pima had the lowest (0.710 ± 0.02). Gene diversity for the Pima was significantly lower than all other populations ($P <$ 0.02) except the Cheyenne. Interestingly, variance in microsatellite repeat length was not correlated with standard gene diversity estimates. Both the Pima and Euro-American samples had quite similar values of repeat length variance (Table 1) yet considerably different gene diversity estimates.

Average gene diversity estimates for each of the populations were similar to those found in a previous study using microsatellite markers on chromosome 20 (Urbanek et al., 1996). The Finnish population falls within the range of gene diversity seen throughout Europe and does not exhibit reduced heterozygosity.

mtDNA diversity

Mitochondrial DNA diversity is shown in Table 2. No significant reduction in mtDNA diversity was found for the Finns when compared to other Europeans. In fact, the Finnish population is quite similar to other European populations in haplotype and nucleotide diversity. Although haplotype diversity was low for the Swedes, their level of nucleotide diversity was consistent with other European populations. One hundred and thirty-five distinct mtDNA haplotypes were observed in the five populations. The number of individuals per population possessing each mtDNA haplotype is given in Appendix A. An interesting pattern of mtDNA variation was observed at the continental level. Although mtDNA haplotype diversity was lower in Pima (0.932 ± 0.011) than in Finns (0.964 ± 0.007), Pima and Cheyenne mtDNA nucleotide diversity was almost three times higher than European mtDNA diversity. Most importantly, not a single mtDNA haplotype was shared between the two Native American populations. These patterns of mtDNA nucleotide and haplotype diversity, which have also been observed in other Amerind populations (Stone and Stoneking, 1998), reveal that pairwise sequence differences within Amer-

TABLE 3. Y chromosome diversity in five populations

Population	n	k	h	Sw
Finland	89	29	0.877 ± 0.013	19.4 ± 0.34
Sweden	45	25	0.965 ± 0.008	26.6 ± 0.56
Euroamerican	32	22	0.964 ± 0.005	22.0 ± 0.43
Cheyenne	31	16	0.905 ± 0.019	21.8 ± 0.50
Pima	55	32	0.959 ± 0.007	21.2 ± 0.42
Overall	252	111	0.975 ± 0.002	27.0 ± 0.19

h, haplotype diversity; k, number of haplotypes; n, sample size; Sw, twice the mean variance in microsatellite repeat length.

ind populations are quite large. This suggests that the ancestral founder population was highly polymorphic (Ward et al., 1991) and/or that after colonization of the New World, Native American populations expanded in size rapidly and in relative isolation (Shields et al., 1993).

As seen elsewhere (DiRienzo and Wilson, 1991), the phylogeny of European mtDNA haplotypes exhibits a star topology whose root is the Cambridge Reference Sequence (CRS) (Anderson et al., 1981). This pattern indicates rapid demographic growth and has been confirmed by pairwise sequence comparisons of mtDNA (Rogers and Harpending, 1992; Harpending et al., 1993; Rogers and Jorde, 1995). The CRS has a frequency of about 20% in nearly all European populations. It was found in the Finnish population at a frequency of 23% (haplotype 1 in Appendix A), and the Finnish mtDNA phylogeny (data not shown) was quite similar to those found using other European data sets (see Richards et al., 1996).

Y-chromosome haplotype diversity

One hundred and eleven Y-chromosome haplotype configurations were observed (see Appendix B) in the sampled populations. As shown in Table 3, the Finns possess the lowest haplotype diversity of all populations, 0.877 ± 0.013. Y-haplotype diversity for the other populations ranged from 0.905–0.965. A similar pattern was observed in Finns by Sajantila et al. (1996) using fewer Y-chromosome loci. Significant differences in Y-chromosome diversity exist between the Finns and both European populations ($P < 0.02$). No significant differences in Y-haplotype diversity were observed between any other pairs of populations.

Three Y-chromosome haplotypes were observed in high frequency (more than ten copies), while 94 were observed in less than three copies. One of the three common Y haplotypes (haplotype A/76) was found exclusively in Finns at a frequency of 30%. Another common Y haplotype (haplotype B/102) was found in 15% of Finns and 11% of Swedish males. A third common haplotype (haplotype D/69) was observed only in Pima males. In the two European populations, many of the single-copy haplotypes were only one or two mutational steps away from a common haplotype from within the same population. If we assume that the oldest haplotypes are those that are in high frequency, then we can infer that the closely related rare haplotypes have recently evolved from the common haplotypes from within that population and are not due to gene flow.

Allele frequencies for each Y-chromosome locus are shown in Table 4. The Native American populations possessed less diversity and fewer alleles at all Y loci except one, *DYS392*, where the Pima possess the largest number of alleles. While most observations of microsatellite mutations have not found changes in allele size of more than one or two repeat units (Weber and Wong, 1993), there is a possibility of mutations of larger size changes (DiRienzo et al., 1994). Allele sizes for the *DYS392* locus ranged from 248–257 base pairs in all populations except the Pima. The presence of the 260, 263, and 266 alleles in the Pima most likely is due to a random mutational event of larger effect that may have occurred after the founding of the Americas. Subsequently, these alleles may have risen to high frequencies in the population due to drift. Nonetheless, reduced Y-chromosome diversity of the Pima is reflected in the average variance in microsatellite repeat length (Sw) across all seven loci. The variance is greatest for Swedes and Euro-Americans (Table 3).

Y-chromosome haplotype phylogeny

The phylogeny relating Y-chromosome haplotypes is shown in Figure 1. Groups A and D are two population-specific clades of Y haplotypes observed in Finnish and Pima males, respectively. Groups A and B were

TABLE 4. Allele frequency distributions in five populations

Locus and allele[1]	Finn	Swede	Euro-American	Pima	Cheyenne
DYS388					
126	0	0.031	0	0	0
129	0.775	0.406	0.911	0.873	0.613
132	0.022	0.063	0.022	0.091	0.387
135	0.202	0.500	0.067	0	0
138	0	0	0	0.018	0
141	0	0	0	0.018	0
DYS389a					
243	0	0.031	0	0.018	0
247	0.258	0.531	0.222	0.109	0.032
251	0.292	0.313	0.578	0.481	0.710
255	0.449	0.125	0.200	0.382	0.258
DYS390					
202	0	0.063	0.089	0.073	0.097
206	0.011	0.250	0.089	0.018	0
210	0.360	0.406	0.111	0.182	0.581
214	0.517	0.219	0.533	0.473	0.290
218	0.112	0.063	0.133	0.255	0.032
222	0	0	0.044	0	0
DYS391					
283	0.360	0.719	0.533	0.673	0.839
287	0.629	0.250	0.467	0.327	0.161
291	0.011	0.031	0	0	0
DYS392					
248	0.337	0.594	0.244	0.091	0.323
251	0	0.031	0.156	0	0
254	0.034	0.281	0.422	0.109	0.065
257	0.629	0.094	0.178	0.218	0.613
260	0	0	0	0.291	0
263	0	0	0	0.145	0
266	0	0	0	0.145	0
DYS393					
120	0	0	0	0.055	0.355
124	0.416	0.906	0.889	0.727	0.419
128	0.573	0.094	0.089	0.127	0.129
132	0.011	0	0.022	0.091	0.097
DYS394					
239	0	0.031	0	0	0
243	0.011	0.063	0.178	0.418	0.323
247	0.809	0.656	0.644	0.491	0.290
251	0.101	0.188	0.067	0.055	0.323
255	0.022	0.031	0.111	0	0
259	0.045	0.031	0	0.036	0.065
263	0.011	0	0	0	0
*DYZ3 Hin*dIII site					
+	0.876	0.625	0.311	0.145	0.355
−	0.124	0.375	0.689	0.855	0.645
Number of chromosomes	89	32	45	55	31

[1] For microsatellite loci, allele sizes (bp) are presented; for the *DYZ3* locus, + denotes the presence of the *Hin*dIII site.

observed as two major star-shaped clusters in the phylogeny. This is indicative of a population expansion from two common Y haplotypes. Group A haplotypes coalesced to a common Y haplotype whose frequency was observed at 30% in Finland. The Pima specific (group D) clade was not surprising since the population was known to have experienced at least one population bottleneck. In fact, this sample was previously shown to be relatively homogeneous and quite divergent from European populations (Urbanek et al., 1996). However, the divergent clade of Finnish Y-chromosome haplotypes lends support to the Finnish population bottleneck hypothesis.

The Y-chromosome phylogeny also revealed substantial amounts of haplotype sharing and clades of related haplotypes among Europeans. Group B and C haplotypes were observed in all of the European populations. A majority of the Finnish Y

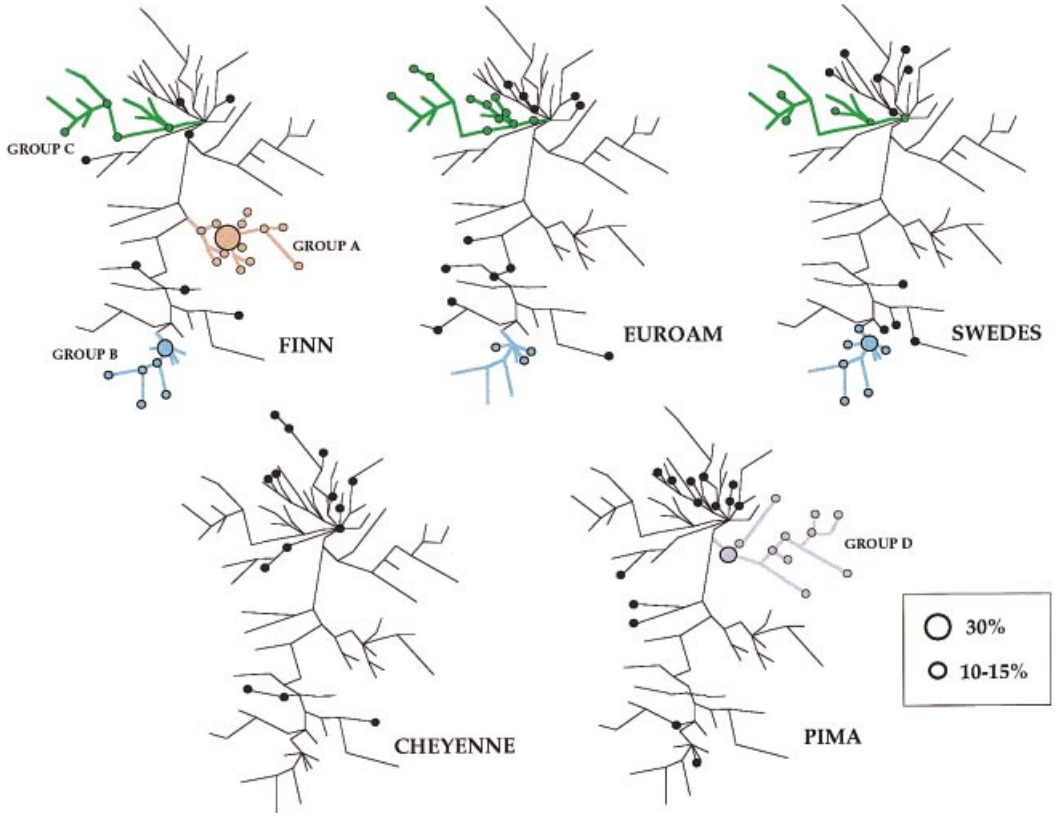

Fig. 1. Y chromosome haplotype phylogeny. Unrooted phylogram relating Y-chromosome haplotypes from five populations. Circles represent observed haplotypes from select populations. The sizes of the circles are proportional to haplotype frequency, and branch lengths are proportional to mutational steps. Group A haplotypes (Finn exclusive) are colored orange, group B haplotypes are blue, group C haplotypes are green, and group D (Pima exclusive) are violet. Phylogeny was inferred using maximum parsimony and represents one of several thousand equally parsimonious trees. Groups A–D were found in 100% of strict consensus trees.

haplotypes were found in groups A and B, while the Euro-American and Swedish Y haplotypes were more evenly distributed throughout the phylogeny (Fig. 1). Euro-American Y-chromosome haplotypes were highly prevalent (50%) in group C.

The majority of Native American Y-chromosome haplotypes clustered separately from Europeans; however, there were four Y haplotypes shared between the Native American and European populations. After careful evaluation of the Native American pedigree and heritage data, we discovered that the four shared haplotypes were due to Euro-American gene flow.

Population differentiation

Genetic variance statistics for the Y-chromosome, autosomal, and mtDNA data are shown in Table 5. AMOVA reveals that Y-chromosome microsatellite diversity is nonrandomly distributed across populations. The Φ_{ST} value for all five populations consisting of a European group (Finns, Swedes, and Euro-American) and a Native American group (Pima, and Cheyenne) is 0.349 ($P < 0.001$). This means that 35% of the total variance of Y-chromosome haplotypes is attributable to differences between populations. The amount of variance attributed to continental differences (Europe/

TABLE 5. Genetic differentiation of populations

Genetic system and type of comparison	All populations		Europeans		Native Americans	
	P value	**Φ** statistic	*P* value	**Φ** statistic	*P* value	**Φ** statistic
Y chromosome						
Among groups	<0.05	**Φ**CT = 0.214		—		—
Among populations/ within groups	<0.001	**Φ**SC = 0.172		—		—
Within populations	<0.001	**Φ**ST = 0.349	<0.001	**Φ**ST = 0.189	0.004	**Φ**ST = 0.137
mtDNA						
Among groups	<0.001	**Φ**CT = 0.144		—		—
Among populations/ within groups	<0.001	**Φ**SC = 0.100		—		—
Within populations	<0.001	**Φ**ST = 0.230	<0.001	**Φ**ST = 0.057	<0.001	**Φ**ST = 0.151

	All populations		Europeans		Native Americans	
	P value	*R* statistic	*P* value	*R* statistic	*P* value	*R* statistic
Y chromosome						
Within populations	<0.02	***R***ST = 0.178	<0.01	***R***ST = 0.090	<0.01	***R***ST = 0.190
Autosomal microsatellite DNA						
Within populations	<0.05	***R***ST = 0.054	<0.01	***R***ST = 0.043	<0.01	***R***ST = 0.080

Americas) was estimated to be 21% ($P < 0.05$). We note that these variance estimates assume that heterozygosity levels are equal for all populations (Urbanek et al., 1996). In order to measure the extent of differentiation within the two continental groups, we applied the AMOVA separately to Europeans and to Native Americans. Significant Y-chromosome differentiation was observed between the three European populations. Table 5 shows that 19% ($\Phi_{ST} = 0.189$, $P < 0.001$) of European Y-chromosome variance was attributed to differences between the Finnish, Swedish, and Euro-American populations. Native American Y-chromosome differentiation exhibited a slightly lower value of 0.137.

While the AMOVA of Y-chromosome variation estimated haplotype differences using all eight Y-chromosome loci including the DYZ3 restriction site, the R_{ST} estimates were calculated using only the variance in repeat lengths of the seven microsatellite loci. The pattern of differentiation observed using the R statistic was quite different than the pattern observed using AMOVA. R_{ST} for all five populations was 0.178; among the Europeans it was 0.090 and among Native Americans 0.190. Autosomal microsatellite differentiation among the five populations was also estimated using the R statistic. R_{ST} for all five populations was 0.054 ($P < 0.05$). Table 5 also shows that R_{ST} estimates for the Y chromosome are two to three times higher than the autosomal estimates, consistent with the fourfold-smaller effective population size for the Y chromosome compared to the autosomes (excluding the X chromosome).

Differentiation between populations for mtDNA control region sequences was not as high as observed for the Y chromosome. AMOVA results for mtDNA revealed that for five populations, consisting of a European group (Finns, Swedes, and Euro-American) and a Native American group (Pima and Cheyenne), about 23% of the total genetic variance is attributable to differences between populations. This variance estimate is consistent with estimates from previous studies using mtDNA (Stenico et al., 1996; Jorde et al. 1995; Stoneking et al. 1990; Merriwether et al., 1991). Much of the variance in mtDNA diversity is attributable to continental differences between Europeans and Native Americans ($\Phi CT = 0.144$, $P < 0.001$). Differences between the three European populations accounts for only 5% ($P <$

0.001) of the total variance (Table 5), while Native American mtDNA differentiation was three times higher, at 15% ($P < 0.001$). The high mtDNA differentiation among Native Americans is mainly due to drift and is concordant with other observations of Native American mtDNA control region sequence variation (Shields et al., 1993; Stone and Stoneking, 1998). The pattern of European mtDNA differentiation is consistent with diversity and phylogenetic analyses of European mtDNA control region sequences which reveal a high degree of homogeneity among Europeans (Richards et al., 1996).

DISCUSSION

Genetic support for a male-specific bottleneck was found from Finnish Y-chromosome microsatellite variation. Data was collected and analyzed from different genetic systems consisting of Y-chromosome microsatellites, mtDNA HVS-1 sequences, and autosomal microsatellites from five populations. Finnish genetic variation was compared to four populations from two continents. Distinct differences in Y-chromosome variation between Finns and other Europeans was observed, yet variation at autosomal and mtDNA loci was indistinguishable among the Europeans. Finnish Y-haplotype diversity was even lower than the diversity observed in two Native American populations which have experienced recent population bottlenecks. While the Cheyenne and Pima represent may represent a restricted Amerindian sample, their patterns of autosomal, mitochondrial, and Y-chromosome variation should be consistent with other Amerindians who experienced similar population histories.

Several features of the R_{ST} results were noteworthy. Higher autosomal gene differentiation was observed among the Native American populations (8%) than among the European populations (4%). This pattern of differentiation reflects unequal divergence between populations and can be attributed to higher levels of genetic drift operating in the two Native American populations. The higher estimates of R_{ST} for the Y chromosome in Native American populations are likely due to many factors, such as differences in effective population size, the number of loci sampled, and the complete linkage of microsatellite loci on the Y chromosome. We note that the performance of R_{ST} is best when the time between population divergence is large and when mutation, not genetic drift, is the dominant force creating differentiation (Slatkin, 1995). These assumptions may not hold true for the Native American populations which have been strongly affected by drift.

With regard to the two measures of Y-chromosome differentiation between European populations, quite distinctive features of Y-chromosome variation were revealed. Slatkin (1995) defined R_{ST} for microsatellite data at a single locus. The measure takes into account differences in microsatellite allelic size. For multilocus microsatellite data, a weighted average across loci is used. Michalakis and Excoffier's (1996) Φ_{ST} is analogous to R_{ST}; however, molecular haplotypic differences and frequencies are utilized in the Φ_{ST} estimate. Thus, the low R_{ST} value revealed that the allelic distributions for Y-microsatellite loci were similar for the European populations; however, the high Φ_{ST} revealed large differences in Y-haplotype configuration between Finns and other Europeans. The opposite pattern was observed for Native American Y-chromosome variation. The high R_{ST} value and slightly lower Φ_{ST} suggest that, although different alleles were observed within the two populations, molecular distances between Y haplotypes were moderate. This pattern is attributable to different ranges of allele sizes for loci such as *DYS392*, where the greatest number of alleles were observed in the Pima.

The discordance between Φ_{ST} estimates for mtDNA and the Y chromosome may be due to greater vulnerability of the Y chromosome to the effects of genetic drift than mtDNA. This may reflect differences in mtDNA and Y chromosome effective population sizes, mainly due to the large variance in male reproductive success and/or differences in male and female dispersal patterns. Another explanation for the differences in Φ_{ST} estimates for mtDNA and the Y chromosome may be differences in mutation rates for the two genetic systems. While the muta-

tion rate is lower for the mtDNA control region (1.4×10^{-5}) (Lundstrom et al., 1992) than microsatellites on the Y chromosome (2.1×10^{-3}) (Heyer et al., 1997), the number of polymorphic sites are greater for the control region. Thus, the potential for higher variation exists for mtDNA haplotypes since each polymorphic nucleotide site within HVS-1 of the mtDNA control region is in essence a locus when compared to the seven microsatellites used to construct Y-chromosome haplotypes.

Reduced number of males

To date, the Finns are the first European population to exhibit a marked reduction of Y chromosome diversity (Table 3). A similar conclusion was drawn from Sajantila et al (1996) using two Y-chromosome microsatellite loci and the Y Alu polymorphism (YAP) locus (Hammer, 1994). Our finding of low Y-chromosome microsatellite diversity for the Finns is quite striking, especially since studies using fewer microsatellite loci have found significantly higher Y-chromosome microsatellite diversity estimates in other European populations (Roewer et al., 1996; Cooper et al., 1996).

Finnish Y-chromosome variation is in strong contrast to mtDNA and autosomal microsatellite data. Estimates of Finnish gene diversity for the two latter genetic systems are consistent with other European populations which have not undergone a population bottleneck. We note that the mutation rate for the control region of mitochondrial DNA is more than 20 times faster than that of nuclear DNA (Greenberg et al., 1983) excluding microsatellites, and nucleotide positions within the control region evolve at different rates (Wakely, 1993). A recent population genetic study on the Finns attempted to deal with the disparate rates at different positions by excluding positions in the analysis which have mutated more than others (Sajantila et al., 1996). They confined the analysis of mitochondrial control region sequences to positions that were slowly evolving in hopes of defining mtDNA lineages that existed before the putative bottleneck. However, using only the slowly evolving positions in the analysis is problematic because variation for all of the European populations, not just the Finns, will be reduced. As more positions are excluded from the analysis, the number of European sequences identical to or one mutational step away from the CRS increases greatly. This is the case because every European population shows a similar pattern of divergence from the CRS and the CRS frequency is similar throughout the continent (Richards et al., 1996).

Genetic heterogeneity of the Finns

As stated earlier, many investigators trace the origin of the Finnish population to an area south of the Gulf of Finland which includes present-day Estonia (Nevanlinna, 1972; Luho, 1976; Fodor and Czeizel, 1991). However, a competing theory contends that the early settlers arrived from two locations, one in this southern region and the second group from the east near the Lake Ladoga region (Norio, 1981; Meinander, 1973; Eriksson, 1973). Archeological evidence appears to support the 'dual origins' theory with considerable cultural differences between eastern and western Finland. These cultural differences, exhibited by eastern and western varieties of plows, sleighs, and architecture (Luho, 1976; Vilkuna, 1976), date back almost 1,000 years (Hajdu, 1975).

Recently, we have shown that two major clades of divergent Y chromosome haplotypes exist within the Finnish population (Kittles et al., 1998). Group A haplotypes were a monophyletic clade of haplotypes not found in other European populations. A second major clade of Y haplotypes, group B, was found in Swedish and Euro-American males. A minimum of ten mutational steps separate the two common Y-chromosome haplotypes (haplotypes A and B). Thus, it is very unlikely that group A and B haplotypes originate from the same source population. The geographic pattern of Y-haplotype variation suggests that haplotype A may have entered Finland from the east (near the Lake Ladoga region), while settlers possessing haplotype B traveled across the Gulf of Finland and the Swedish border (Kittles et al., 1998). Estimated expansion times for the haplotypic groups suggest that group B

haplotypes may have arrived in Finland during the diffusion of agriculturalism through Europe.

The presence of the *DYZ3 Hin*dIII restriction site was used in conjunction with the Y-chromosome microsatellites to increase haplotype resolution in order to more precisely define the major Y lineages. The *DYZ3 Hin*dIII (+) is ancient, most likely predating the expansion of humans out of Africa since it was observed in geographically diverse populations on different haplotype back grounds (Santos et al., 1995; Kittles, unpublished data). The *DYZ3 Hin*dIII restriction site was found on 88% of the Finnish Y chromosomes (groups A and B). The other 12% of Finnish Y chromosomes (group C), being quite divergent (mutationally) and rare, may have arrived in Finland relatively recently.

The Y-chromosome data presented in this study have helped resolve the question of a Finnish population bottleneck. Evidence for distinct Y-chromosome lineages and low Y-chromosome diversity in Finland provides support for at least two bottleneck events affecting males. In fact, multiple founder effects best explain the high incidence and geographic pattern of genetic diseases within Finland. This is especially true since a majority of the genetic diseases result from one founder mutation, revealing major founder effects within the Finnish population.

The disparate pattern of diversity observed for Finnish mtDNA and the Y chromosomes suggests that females may have been incorporated from other European populations, while males were restricted. However, we note that caution should be used when comparing haplotype diversity estimates for the two genetic systems. Diversity within Finland for autosomal microsatellite loci is similar to European populations which have not undergone a population bottleneck. This may be due to the high mutation rate of microsatellites. However, the same phenomenon was not seen in the Native Americans which may have undergone several population bottlenecks throughout history. This being the case, other factors such as sex-biased gene flow and recent internal migration between Finnish subisolates may be responsible for high gene diversity in Finns. Further geographic sampling within Finland will be needed in order to clarify this. Excluding the Y-chromosome data, genetic variation in Finland does not reveal evidence of reduced heterozygosity. This suggests that there may be a detectable amount of underlying autosomal homozygosity within certain regions formed during the initial settlement of Finland. This pattern may also have been diluted by recent immigration of females from other European populations. While this gene flow would affect autosomal (including X chromosome) and mtDNA gene diversity, it would not affect variation at the paternally transmitted Y chromosome. Consequently, Y-chromosome analysis provides an independent approach to test the isolate hypothesis for the Finnish population.

ACKNOWLEDGMENTS

We thank Drs. Jaakko Lappalainen, S.O.Y. Keita, Markus Perola, and Leena Peltonen for helpful discussions. We also appreciate technical support from Elisa Moore and Longina Akhtar. An earlier draft of this paper was improved by comments from two anonymous reviewers.

APPENDIX A. Sequence variants in 135 mtDNA haplotypes from five populations[1]

Haplotype	Variants						Population				
		11111111111111111111	11111111111112222222	22222222222222222222	22222222222222333333	3333333333					
	122223344455678889	00012222233344445555	66666667899990000011	11222333333444556666	67777778899999000011	2223333345			Euro-		
Haplotype	912580746712901581	13671245903906793789	01234690102470157912	79356234578037154678	90123581834567124527	1782345937	Finn	Swede	American	Cheyenne	Pima
1	GTTGACGCCAATTTCCCC	TTGTTGCCTGGAACATCAAA	ACCCCTCCGGATACTCACCA	ACACTGCCACCTCCGCCACC	CACCCTCTTAAAGCTTCCAA	CAACTTCTCC	13	2	8	.	.
2	....	....	....	....	.....C..............	....	.	11	2	.	.
3	....	....	..............C.....	....	....	....	.	.	1	.	.
4	....	....	....	....	.......C............	....	2	.	1	.	2
5	....	....	....	......T.............	....	....	1	.	.	.	.
6	.....T............	....	....	....	....	....	1	.	.	.	.
7	....	...........G........	....	....	....	....	3	.	.	.	.
8	....	......T.............	....	....	....	....	1	.	.	.	.
9	....	....	....	....	....	....C.....	1	.	.	.	.
10	....	....	....	..............A.....	....	....	1	1	.	.	.
11	....	....	....	....	....	.......C..	1	.	.	.	.
12	.........G........	....	....	....	....	....	1	.	.	.	.
13	....	........C...........	....	....	....	....	.	.	.	.	.
14	....	....	....	......T......T......	....	....	1	.	.	.	.
15	....G.............	....	....	......T......T......	....	....	1	.	.	.	.
16	....	....	..........T......T..	....	....	..T.......	1	.	.	.	.
17	....	....	....	....	....	T...C..C..	1	.	.	.	.
18	.............C....	....	....	......T......T.....T	....	....	1	.	.	.	.
19	.............C....	....	......T.............	......T......T.....T	....	....	1	.	.	.	.
20	.................A	....	......T.............	......T......T......	..T.................	....	1	.	.	.	.
21	....	....	......T.............	......T......T......	....	.......C..	1	2	.	.	.
22	....	...........G........	....	......T......T......	....	....	1	.	.	.	.
23	....	....	....	......T......T......	..T.................	....	1	.	.	.	.
24	....	.........A..........	....	....	.....C..............	....	.	.	.	.	.
25	....	....	G...................	....	.....C..............	....	2	.	.	.	.
26	T.................	....	....	....	.....C..............	....	1	.	.	.	.
27	...........C......	....	....	G...................	.....C..............	....	.	.	1	.	.
28	.......T..........	....	....	...................T	..............C.....	....	.	.	1	.	.
29	....	....	....	.......T............	....	....	.	.	1	.	.
30	....	....	....	...................T	....	....	1	.	.	.	.
31	....	..A.................	....	....	....	....	1	.	1	.	.
32	.............C....	..A.................	....	....	....	....	.	.	1	.	.
33	....	..A.................	............C.......	....	....	....	.	.	1	.	.
34	....	..A.................	....	....	...........T........	....	.	1	.	.	.
35	.............C....	..A.................	....	....	.........G..........	....	2	.	.	.	.
36	....	....	....	....	..T.T..C............	....	.	.	1	.	.
37	....	....	....	.............T......	....T..........C....	....	.	.	1	.	.
38	....	....	....	.................G..	....	....	.	.	1	.	.
39	....	.C..................	....	....	..T.................	....	.	.	1	.	.
40	........T.........	.C..................	....	....	....	....	3	.	.	.	.
41	....	.C..................	....	....	..T...........C.....	....	1	.	.	.	.
42	........T........G	.C..................	....	....	....	....	.	.	1	.	.
43	........T.........	.C..................	.......T............	....	....	....	1	.	.	.	.
44	........T.........	.C..................	.......T............	...............T....	....	....	1	.	.	.	.
45	..A...............	.C..................	....	....	..T.T..C............	....	1	.	.	.	.
46	....	.C.......A..........	....	....	..T.T...............	....	.	.	1	.	.
47	....	.C..................	....	..GA................	..T.T..C............	....	2	.	.	.	.

(*continued*)

Haplotype	Variants						Population				
		11111111111111111111	11111111111112222222	22222222222222222222	22222222222222333333	3333333333					
	122223344455678889	00012222233344445555	66666667899990000011	11222333333444556666	67777778899999000011	2223333345			Euro-		
Haplotype	912580746712901581	13671245903906793789	01234690102470157912	79356234578037154678	90123581834567124527	1782345937	Finn	Swede	American	Cheyenne	Pima
48	....	.C..........G.......	...T.C..............	....	..T.................	....	1	.	.	.	.
49	................T.	....	C....C.....C........	....	....	....	.	.	.	1	.
50	................T.	C...................	C....C.....C........	....	....	....	.	.	.	.	1
51	................T.	....	C....C.....C.T......	....	....	......T...	.	.	.	1	.
52	.A..............T.	C...................	C....C.....C........	.............T......	....	....	.	.	.	.	1
53	............C.....	....	C.T..C.....C........	....	...............CT...	....	.	.	.	.	1
54	...A..........T...	..................GC	C....C.....C........	....	....	....	.	.	.	1	.
55	...A..........T...	..............G...GC	C....C.....C........	.............T......	....	.........A	.	.	.	1	.
56	....	....	C....C.....C...G....	.G........T.........	....	....	.	.	.	.	1
57	....	....	C....C.....C........	..........T..T......	....	.........A	.	.	.	.	1
58	....	....	C....C.....C........	..............A.....	....	......G.G.	.	.	.	1	.
59	....	....	C....C.....C........	..........T.........	....	.........A	.	.	.	.	1
60	....	....	C....C.....C........	..........T.........	....	....	.	.	.	.	8
61	....	....	C....C.....C........	..............A.....	.................T..	....	.	.	.	.	.
62	............C.....	.................—..	...T.C.....C........	....	....	....	.	.	.	.	2
63	....	....C...............	.....C..............	...G.........T......	....	....	.	1	.	.	.
64	....	....C...............	.....C..............	.............T......	..................G.	....	.	1	.	.	.
65	....	.................—..	...T.C.....C........	....	....	....	.	.	.	.	1
66	....	....C...............	.....C..............	.............T......	....	....	.	3	.	.	.
67	....	....	.....C..............	....	....	....	.	.	1	.	.
68	............C.....	.................—..	...T.C.....C........	.............A..G...	..................T.	....	.	.	.	.	1
69	....	....C..T............	.....C..............	.............T......	..................G.	....	.	2	.	.	.
70	....	....C..T............	.....C..............	.............T......	....	....	.	1	.	.	.
71	....	....C...............	.....C..............	.............T......	......T.............	....	1	.	.	.	.
72	....	....	.....C.........G....	.............T......	..........G.........	....	1	.	.	.	.
73	..........G.......	....	.....CT.............	.............T......	....	....	1	.	.	.	.
74	....	.............A......	......T.............	.............T......	....	....	.	.	1	.	.
75	....	....	.............T......	....	..T............C....	.......C..	.	.	.	.	1
76	....	....	C....C..............	.............T......	....	....	1	.	.	.	.
77	....	....	.....CT.............	.............T......	....	....	4	.	.	.	.
78	....	....	.....C..............	....	.....C..............	....	.	.	1	.	.
79	....	....	.T...........T......	..........T.......T.	............A.......	.......C..	.	.	.	1	.
80	....	.CA.................	....T........T......	..................T.	............A.......	.......C..	.	.	.	1	.
81	....	.CA.................	.............T......	..................T.	............A.C.....	.......C..	.	.	.	1	.
82	....	....	C....C.......T......	....C.............T.	............A.......	.......C..	.	.	.	1	.
83	................T.	....	.............T......	..................T.	............A.......	.......C..	.	.	.	4	.
84	................T.	....	.............T..G...	..................T.	............A.......	.......C..	.	.	.	1	.
85	................T.	....	.............T......	..................T.	............A..C....	.......C..	.	.	.	3	.
86	................T.	....	......T......T......	..................T.	............A.......	.......C..	.	.	.	3	.
87	................T.	....	......T......T......	..................T.	............A.......	.......C.T	.	.	.	1	.
88	................T.	....	......T......T......	....C.............T.	............A.......	.......C..	.	.	.	2	.
89	................T.	....	......T......T......	..................T.	............A......G	....	.	.	.	1	.
90	................T.	.................T..	.............T......	.......T...A......T.	............A.......	.......C..	.	.	.	.	2
91	....	....	.............T......	....	..T..C.........CT...	.........A	.	.	.	1	.
92	....	....	.............T......	....	..T..C..C......CT...	....	.	.	.	1	.
93	....	....	.............T......	....	..T..C.........CT...	....	.	.	.	3	.
94	....	....	.............T......	....	.....C.........CT...	....	.	.	.	.	6

95	.....	.....	.....T.....	.....	.....C.....T...	.....	.	.	.	1	.
96	.....C.....	.....T.....	....C.....	.....C.....CT...	.....	.	.	.	.	1
97	....G.....	.....	.....T....T.	.....	.....C.....CT...	.....	.	.	.	1	.
98	.....	.....	.....T....T.	.....	...T.C.....CT...	.....	.	.	.	.	5
99	.....A.....	.....	.....T....T.	.....	...T.C.....CT...	.....	.	.	.	.	1
100	.....A.....	.....T.....TC.....	.....T.....	.....C.....	.....	.	1	.	.	.
101	.....	.....	.....T.....	...T.....	.....C.....CT...	.....	.	.	.	.	1
102	.....C.....	.....	.....T.....	.....T.....	.....C.....	.....	1	.	.	.	.
103	.....	.....	.....C.....T.....	.....	.....	.....	.	.	1	.	.
104	.....	.....	.....T.....	.....	T.....	.....	.	.	1	.	.
105	.....C.....	.....	C....C...A...T.....	.....T.....	.....	...G.CT...	.	.	.	1	.
106	......A.....T.	.....T.....C	C....C.....	.....T.....	.....	.....	.	.	.	.	1
107	.....	.....	C....C.....T.....	.....T...T	.....A.....	.....C....	.	.	.	3	.
108	.....C.....	.....	C....C...A...T.....	.....T.....	.....	.....	.	.	.	1	.
109	.....	.....	C....C.....T.....	.....T...T	.....A.....	.....	.	.	.	2	.
110	.....T..	.....	.....C.....T.....	.....A.....T.....	.....	.....	1	.	.	.	.
111	.....	.....	......T......T.....	.....	T.....C.....	.....	1	.	.	.	.
112	.....	.....	.....T....T.	.....	T.....	.....	1	.	.	.	.
113	.....	.....	.....T.....	.T.....	T.....	.....	1	.	.	.	.
114	.....	.....A.....G...	.....T.....	.....	.....	.....	1	.	.	.	.
115	.....	.....	.....G	.....	.....AT.....	.....	.	.	1	.	.
116	.....C.....	.....	.....A.....	.....	.....	.....	.	.	.	1	.
117	.....C.....	.....	.....	.....	.....C.....	.....	.	.	1	.	.
118	.....C.....	.....	.....C.....	.....	.....	.....	1	.	.	.	.
119	.....C.....	.....	.....	.....	.....	.....	1	.	1	.	.
120	.....	.....	.....	.....	.....C.....	.....	1	1	3	.	.
121	.....	.....C.....	.....	.....	.....C.....	.....	.	.	1	.	.
122	.....	.....	.....C.....	.....	.....C.....	.....	1	.	.	.	.
123	.....	.....	.....	.....T.....	.....C.....	.....	.	.	2	.	.
124	.....T...	.....	.....C.....	.....	.....C.....	.....	.	.	1	.	.
125	.....	.....	.....C.....	.....	.....C.....	.....C..	.	.	1	.	.
126	.....	.....C.....	.....	.....	.....CC.....	.....	.	.	.	.	1
127	.....	.....	.....	.....G.....	.....C.....C....	.....	1	.	.	.	.
128	.....C.....	.....	.....C.....	.....	.....C.....	.....	.	.	1	.	.
129	.....C.....	.....	.....	.....	.....C.....	.....C..	.	.	1	.	.
130	.....C.....	.....	.....	.....T.....	.....C.....	.....	.	.	1	.	.
131	.....	.....	.....	.....T.....	.G.....C.....	.....	1	.	.	.	.
132	.....	.....	.....	.....T.....	.G.....C.....	.T.....	1	.	.	.	.
133	.....	.....	.....	.....T.....	.....C.....	.....	1	.	.	.	.
134	.....	.....	.....TT.....	.....T.....	.....C.....	.....	.	1	.	.	.
135	.....C.....	.....	.....T.....	.....	.....C.....CT...	.....	.	.	.	.	1
							73	28	44	39	40

[1] Variable nucleotide positions (108) from mitochondrial control region sequence (HVS-1) data for 135 haplotypes found in five populations. Position 19 corresponds to position 16043 of the published human reference sequence (Anderson et al. 1981).

APPENDIX B. Y chromosome haplotypes for five populations[1]

Haplotype	*DYS388*	*DYS389*	*DYS390*	*DYS391*	*DYS392*	*DYS393*	*DYS394*	*DYZ3*	Population
1	126	251	218	283	248	124	255	−	S
2	129	243	218	287	263	128	243	−	P
3	129	247	202	283	248	128	251	+	E
4	129	247	202	283	248	132	251	+	P
5	129	247	206	283	257	124	243	−	S
6	129	247	210	283	257	120	247	−	P
7	129	247	210	283	257	124	243	−	S
8	129	247	210	283	248	124	247	+	F
9	129	247	210	283	248	124	251	+	S,F
10	129	247	214	283	254	124	247	−	E
11	129	247	214	283	248	128	243	+	E
12	129	247	218	283	257	128	247	−	C
13	129	247	218	283	263	128	247	−	P
14	129	247	214	287	254	124	247	−	S
15	129	247	218	287	257	128	247	+	F
16	129	251	202	283	248	124	255	+	E
17	129	251	206	283	254	124	247	−	E
18	129	251	206	291	254	124	247	−	S
19	129	251	210	283	248	120	247	+	C
20	129	251	210	283	248	124	243	+	E
21	129	251	210	283	251	124	243	+	E
22	129	251	210	283	254	124	247	−	P
23	129	251	210	283	257	124	247	−	P
24	129	251	210	283	257	128	247	−	C
25	129	251	210	287	254	124	239	−	S
26	129	251	210	287	254	124	247	−	P
27	129	251	210	287	257	124	243	−	P,C
28	129	251	210	287	257	124	259	−	C
29	129	251	210	287	257	128	247	+	F
30	129	251	214	283	248	124	243	+	F
31	129	251	214	283	248	124	247	−	F
32	129	251	214	283	254	124	243	−	P,E
33	129	251	214	283	254	124	247	−	E,S,F
34	129	251	214	283	257	120	247	−	C
35	129	251	214	283	257	124	243	−	P,C
36	129	251	214	283	257	124	247	−	C
37	129	251	214	283	257	124	259	−	C
38	129	251	214	283	260	128	243	−	P
39	129	251	214	283	263	128	243	−	P
40	129	251	214	287	248	124	247	−	E
41	129	251	214	287	254	124	243	−	P
42	129	251	214	287	254	124	247	+	P,F,E,S
43	129	251	214	287	254	128	247	−	E
44	129	251	214	287	257	124	247	−	E
45	129	251	214	287	257	128	247	+	F
46	129	251	214	287	257	132	247	+	F
47	129	251	214	287	260	128	243	−	P
48	129	251	214	287	266	124	243	−	P
49	129	251	218	283	257	120	247	−	P
50	129	251	218	283	257	124	243	−	P
51	129	251	218	283	257	124	247	−	P
52	129	251	218	283	248	124	251	−	F
53	129	251	218	283	248	124	259	−	S
54	129	251	218	287	266	124	243	−	P
55	129	251	218	287	257	124	247	−	E
56	129	251	218	287	248	124	255	−	F
57	129	251	218	287	248	124	259	−	F
58	129	251	218	287	257	128	259	+	F
59	129	251	222	287	251	124	255	−	E
60	129	255	202	283	254	124	243	+	E
61	129	255	202	283	257	132	243	−	C
62	129	255	210	283	257	124	243	−	C
63	129	255	210	283	257	124	247	+	P
64	129	255	210	287	257	124	247	+	F
65	129	255	210	287	248	128	247	+	F
66	129	255	210	287	257	128	247	−	F
67	129	255	214	283	254	124	247	−	E
68	129	255	214	283	257	124	247	−	F
D/69	129	255	214	283	260	124	247	−	P

(*continued*)

APPENDIX B. (continued)

Haplotype	*DYS388*	*DYS389*	*DYS390*	*DYS391*	*DYS392*	*DYS393*	*DYS394*	*DYZ3*	Population
70	129	255	214	283	260	124	259	+	P
71	129	255	214	283	257	128	251	+	C,F
72	129	255	214	287	263	124	243	−	P
73	129	255	214	287	254	124	247	−	E,C
74	129	255	214	287	263	124	259	−	P
75	129	255	214	287	254	128	247	−	F
A/76	129	255	214	287	257	128	247	+	F
77	129	255	214	287	257	128	251	+	F
78	129	255	214	287	257	128	255	+	F
79	129	255	214	287	257	128	263	−	F
80	129	255	214	291	257	128	247	+	F
81	129	255	218	283	263	124	243	+	P
82	129	255	218	283	251	124	251	−	E
83	129	255	218	287	254	124	247	−	E
84	129	255	218	287	257	124	247	−	E
85	129	255	218	287	251	124	251	−	E
86	129	255	218	287	257	128	247	+	F
87	132	247	206	283	248	124	247	+	S
88	132	251	202	283	248	124	247	+	P
89	132	251	210	283	248	120	251	+	C
90	132	251	210	287	254	124	247	−	P
91	132	251	210	287	257	124	247	+	F
92	132	251	214	287	254	124	247	−	C
93	132	255	202	283	257	132	243	−	C
94	132	255	210	283	251	128	251	+	S
95	132	255	210	283	254	132	255	+	E
96	132	255	214	283	260	124	247	−	P
97	132	255	218	283	260	132	243	−	P
98	135	247	202	283	248	124	247	+	S
99	135	243	206	287	248	124	251	+	S
100	135	247	206	283	248	124	251	+	F
101	135	247	206	283	248	124	247	+	E,S
B/102	135	247	210	283	248	124	247	+	S,F
103	135	247	210	283	248	124	251	+	S,F
104	135	247	210	283	248	124	259	+	F
105	135	247	206	283	251	124	247	+	E
106	135	247	210	283	254	124	251	+	S
107	135	251	206	283	248	124	247	+	S
108	135	255	210	283	248	124	247	+	S
109	138	247	206	283	248	124	247	+	P
110	141	255	214	283	260	124	247	−	P
111	129	255	210	283	257	128	247	−	P

[1] Allele sizes for Y-chromosome microsatellite loci and *DYZ3* allele designation for 111 Y-chromosome haplotypes found in five populations. C, Cheyenne; E, Euro-American; F, Finn; P, Pima; S, Swede.

LITERATURE CITED

Anderson S, Bankier A, Barrell B, de Bruijn M, Coulson A, Drouin J, Eperon I, Nierlich D, Roe B, Young I. 1981. Sequence and organization of the human mitochondrial genome. Nature 290:457–465.

Bertranpetit J, Sala J, Calafell F, Underhill P, Moral P, Comas D. 1995. Human mitochondrial DNA variation and the origin of Basques. Ann Hum Genet 59:63–81.

Bowcock A, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd J, Cavalli-Sforza L. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457.

Cavalli-Sforza LL, Piazza A. 1993. Human genomic diversity in Europe: a summary of recent research and prospects for the future. Eur J Hum Genet 1:3–18.

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

Deka R, Shriver M, Yu L, DeCroo S, Hundrieser J, Bunker C, Ferrell R, Chakraborty R. 1995. Population genetics of dinucleotide (dC-dA)n · (dG-dT)n polymorphisms in world populations. Am J Hum Genet 56:461–474.

Deka R, Jin L, Shriver M, Yu L, Saha N, Barrabtes R, Chakraborty R, Ferrell R. 1996. Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. Genome Research 6:1177–1184.

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

DiRienzo A, Wilson A. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. Proc Natl Acad Sci U S A 88:1597–1601.

DiRienzo 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 U S A 91:3166–3170.

Donnelly P. 1996. Interpreting genetic variability: the effects of shared evolutionary history. In: Chadwick D, Cardew G, editors. Variation in the human genome. Chichester: John Wiley & Sons. p 25–50.

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

Excoffier L, Smouse P, Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: 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, editors. Genetics of the Hungarian population. Berlin: Springer-Verlag.

Goldstein D, Ruiz-Linares A, Cavalli-Sforza LL, Feldman M. 1995. Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci U S A 92:6723–6727.

Graven L, Passarino G, Semino O, Boursot P, Santachiara-Benerecetti S, Langaney A, Excoffier L. 1995. Evolutionary correlation between control region sequence and restriction polymorphisms in the mitochondrial genome of a large Senegalese Mandenka sample. Mol Biol Evol 12:334–345.

Greenberg B, Newbold J, Sugino A. 1983. Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. Gene 21:33–49.

Greenberg JH. 1987. Language in the Americas. Stanford: Stanford University Press.

Griffiths R, Tavare S. 1994. Ancestral inference in population genetics. Stat Sci 9:307–319.

Guglielmino CR, Piazza A, Menozzi P, Cavalli-Sforza LL. 1990. Uralic genes in Europe. Am J Phys Anthropol 83:57–68.

Hajdu P. 1975. Finno-Ugrian languages and peoples. London: Andre Deutsch Ltd.

Hammer M. 1994. A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761.

Hammer M, Spurdle A, Karafet T, Bonner M, Wood E, Novelletto A, Malaspina P, Mitchell R, Horai S, Jenkins T, Zegura S. 1997. The geographic distribution of human Y chromosome variation. Genetics 145:787–805.

Harpending H, Sherry S, Rogers A, Stoneking M. 1993. The genetic structure of ancient human populations. Current Anthropology 34:483–496.

Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E. 1992. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. Nature Genetics 2:204–211.

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.

Hoglund P, Sistonen P, Norio R, Holmberg C, Dimberg A, Gustavson K, de la Chapelle A, Kere J. 1995. Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. Am J Hum Genet 57:95–102.

Hudson R. 1990. Gene genealogies and the coalescent process. Oxford Surveys in Evolutionary Biology 7: 1–44.

Jobling M, Tyler-Smith C. 1995. Fathers and sons: the Y chromosome and human evolution. Trends Genet 11:449–456.

Jorde L, Bamshad M, Watkins W, Zenger R, Fraley A, Krakowiak P, Carpenter K, Soodyall H, Jenkins T, Rogers A. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am J Hum Genet 57:523–538.

Kimura M, Crow JF. 1964. The number of alleles that can be maintained in a finite population. Genetics 49:725–735

Kimura M, Ohta T. 1978. Stepwise mutation model and the distribution of allelic frequencies in a finite population. Proc Natl Acad Sci U S A 75:2868–2872.

Kittles R, Urbanek M, Goldman D, Long J. 1996. Microsatellite polymorphisms reveal average heterozygosity for Finns comparable with other European populations. Am J Phys Anthropol 22(Suppl):139.

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC. 1998. Dual origins of Finns revealed by Y chromosome haplotype variation. Am J Hum Genet 62:1171–1179.

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.

Long J, Ross M. 1995. BioAutoGraph, version 1.0. Bethesda, MD: Section on Population Genetics and Linkage, Laboratory of Neurogenetics, National Institutes of Alcohol and Alcohol Abuse, NIH, (distributed by authors).

Luho V. 1976. The population and prehistory of Finland. In: Hajdu P, editor. Ancient cultures of the Uralian peoples. Budapest: Corvina Press. p 115–132.

Lundstrom R, Tavare S, Ward R. 1992. Estimating substitution rates from molecular data using the coalescent. Proc Natl Acad Sci U S A 89:5961–5965.

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 7. Helsinki: University of Helsinki Archaeology Institute.

Merriwether DA, Clark A, Balliner S, Schurr T, Soodyall H, Jenkins T, Sherry S, Wallace D. 1991. The structure of human mitochondrial DNA variation. J Hum Evol 33:543–555.

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.

Mountain J, Herbert J, Bhattacharyya S, Underhill P, Ottolenghi C, Gadgil M, Cavalli-Sforza LL. 1995. Demographic history of India and mtDNA-sequence diversity. Am J Hum Genet 56:979–992.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nei M, Roychoudhury K. 1993. Evolutionary relationships of human populations on a global scale. Mol Biol Evol 10:927–943.

Nei M, Takezaki N. 1996. The root of the phylogenetic tree of human populations. Mol Biol Evol 13:170–177.

Nei M, Maruyama T, Chakraborty R. 1975. The bottleneck effect and genetic variability in populations. Evolution 29:1–10.

Nevanlinna HR. 1972. The Finnish population structure: a genetic and genealogical study. Hereditas 71:195–236.

Norio R. 1981. Diseases in Finland and Scandinavia. In: Rothschild H, editor. Biocultural aspects of disease. New York: Academic Press. p 359–415.

Peltonen L, Pekkarinen P, Aaltonen J. 1995. Messages from an isolate: lessons from the Finnish gene pool. Biol Chem Hoppe Seyler 376:697–704.

Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H, Sykes B. 1996. Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203.

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Kniff P. 1996. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum Mol Genet 5:1029–1033.

Rogers A, Harpending H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol Biol Evol 9:552–569.
Rogers A, Jorde L. 1995. Genetic evidence on modern human origins. Hum Biol 67:1–36.
Ruhlen M. 1987. A guide to the world's languages, vol 1. Stanford: Stanford University Press.
Sajantila A, Budowle B, Strom M, Johnson V, Lukka M, Peltonen L, Ehnholm C. 1992. PCR Amplification of alleles at the D1S80 locus: comparison of a Finnish and a North American Caucasian population sample, and forensic casework evaluation. Am J Hum Genet 50:816–825.
Sajantila A, Lahermo P, Anttinen T, Lukka M, Sistonen P, Savontaus M, Aula P, Beckman L, Tranebjaerg L, Gedde-Dahl T, Issel-Tarver L, DiRienso A, Paabo S. 1995. Genes and languages in Europe: an analysis of mitochondrial lineages. Genome Research 5:42–52.
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 U S A 93:12035–12039.
Sanchez-Mazas A, Butler-Brunner E, Excoffier L, Ghanem N, Ben Salem M, Breguct G, Dard P, Lefranc G, Butler R. 1994. New data for Ag haplotype frequencies in Caucasoid populations and selective neutrality of the Ag polymorphism. Hum Biol 66:27–48.
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.
Santos F, Gerelsaikhan T, Munkhtuja B, Oyunsuren T, Epplen J, Pena S. 1996. Geographic differences in allele frequencies of the human Y-linked tetranucleotide polymorphism *DYS19*. Hum Genet 97:309–313.
Shields G, Schmiechen A, Frazier B, Redd A, Voevoda M, Reed J, Ward R. 1993. mtDNA sequences suggest a recent evolutionary divergence for Beringian and Northern North American populations. Am J Hum Genet 53:549–562.
Shriver M, Jin L, Chakraborty R, Boerwinkle E. 1993. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993.
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.
Sokal R, Oden N, Thomson B. 1988. Genetic changes across language boundaries in Europe. Am J Phys Anthropol 76:337–361.
Sokal R, Oden N, Legendre P, Fortin M, Kim J, Vaudor A. 1989. Genetic differences among language families in Europe. Am J Phys Anthropol 79:489–502.
Stenico M, Nigro L, Bertorelle G, Calafell F, Capitanio M, Corrain C, Barbujani G. 1996. High mitochondrial sequence diversity in linguistic isolates of the Alps. Am J Hum Genet 59:1363–1375.
Stone AC, Stoneking M. 1998. mtDNA analysis of a prehistoric Oneota population: implications for the peopling of the New World. Am J Hum Genet 62:1153–1170.
Stoneking M, Jorde L, Bhatia K, Wilson A. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. Genetics 124:717–733.
Swofford DL. 1993 PAUP, version 3.1.1. Washington, DC: Smithsonian Institution.
Templeton A. 1993. The 'Eve' hypothesis: a genetic critique and reanalysis. American Anthropologist 95: 51–72.
Underhill P, Jin L, Lin A, Mehdi S, Jenkins T, Vollrath D, Davis R, Cavalli-Sforza L, Oefner P. 1997. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Research 7:996–1005.
Urbanek M, Goldman D, Long JC. 1996. The apportionment of dinucleotide repeat diversity in Native Americans and Europeans: a new approach to measuring gene identity reveals asymmetric patterns of divergence. Mol Biol Evol 13:943–953.
Valdes A, Slatkin M, Freimer N. 1993. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749.
Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Jarvela I. 1996. The age of human mutation: genealogical and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet 58:506–512.
Vigilant L, Pennington R, Harpending H, Kocher T, Wilson A. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc Natl Acad Sci U S A 86:9350–9354.
Vilkuna K. 1976. The formation of the Finnish peasant way of life. In: Hajdu P, editor. Ancient cultures of the Uralian peoples. Budapest: Corvina Press. p 183–190.
Wakely J. 1993. Substitution rate variation among sites in hypervariable region 1 of human mitochondrial DNA. J Mol Evol 37:613–623.
Ward R, Frazier B, Dew-Jager K, Paabo S. 1991. Extensive mitochondrial diversity within a single Amerindian tribe. Proc Natl Acad Sci U S A 88:8720–8724.
Weber J, Wong C. 1993. Mutation and human short tandem repeats. Hum Mol Genet 2:1123–1128.
Wright S. 1951. The genetical structure of populations. Annuals of Eugenics 1:323–334.
Zerjal T, Dashnyam B, Pandya A, 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183.