European Journal of Human Genetics (1999) 7, 447–458

http://www.stockton-press.co.uk/ejhg

ARTICLE

Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami

Päivi Lahermo[1], Marja-Liisa Savontaus[1,2], Pertti Sistonen[3], Judit Béres[4], Peter de Knijff[5], Pertti Aula[1] and Antti Sajantila[6]

[1]*Department of Medical Genetics, University of Turku, Finland*
[2]*Department of Genetics, University of Turku, Finland*
[3]*Finnish Red Cross, Blood Transfusion Service, Helsinki, Finland*
[4]*'Bela Johan' National Institute of Health, Budapest, Hungary*
[5]*MGC-Department of Human Genetics, University of Leiden, The Netherlands*
[6]*Department of Forensic Medicine, University of Helsinki, Finland*

Y chromosomal polymorphisms were studied in 502 males from 16 Eurasian ethnic groups including the Finns, Saami (Inari Lake area and Skolt Saami), Karelians, Mari, Mokshas, Erzas, Hungarians (Budapest area and Csángós), Khanty, Mansi, Yakuts, Koryaks, Nivkhs, Mongolians, and Latvians. The samples were analysed for polymorphisms in the Y chromosome specific Alu insertion (YAP) and six microsatellites (DYS19, DYS389-I and II, DYS390, DYS392, DYS393). The populations were also screened for the recently described Tat polymorphism. The incidence of YAP$^+$ type was highest in the Csángós and in other Hungarians (37.5% and 17.5%, respectively). In the Karelians and the Latvians it was present at approximately the same level as commonly found in other European populations, whilst absent in our further samples of Eurasian populations, including the Finns and the Saami. Aside from the Hungarians, the C allele of the Tat polymorphism was common in all the Finno-Ugric speaking populations (from 8.2% to 63.2%), with highest incidence in the Ob-Ugrian Khanty. The C allele was also found in the Latvians (29.4%). The haplotypes found associated with the Tat C allele showed consistently lower density than those associated with the T allele, indicating that the T allele is the original form. The computation of the age of the Tat C suggested that the mutation might be a relatively recent event giving a maximum likelihood estimate of 4440 years (95% confidence interval about 3140–6200 years). The distribution patterns of the 222 haplotypes found varied considerably among the populations. In the Finns a majority of the haplotypes could be assigned to two distinct groups, one of which harboured the C allele of the Tat polymorphism, indicating dichotomous primary source of genetic variation among Finnish males. The presence of a bottleneck or founding effect in the male lineages of some of the populations, namely in the Finns and the Saami, would appear to be one likely interpretation for these findings.

Keywords: Y chromosome; polymorphisms; microsatellites; Yap; Tat; Finno-Ugric

Correspondence (present address): Päivi Lahermo, University of Helsinki, Finnish Genome Center, PO Box 21 (Tukholmankatu 2), FIN-00014 University of Helsinki, Finland. Tel: +358 9 191 26754; Fax: +358 9 191 26789; E-mail: paivi.lahermo@helsinki.fi
Received 3 April 1998; revised 13 November 1998; accepted 16 December 1998

Introduction

Recent studies using autosomal markers and mtDNA (mtDNA) polymorphisms indicate that the Finns, who speak a Finno-Ugric language, share a common mtDNA gene pool with the Indo-European language speakers in Europe, which argues against their status as genetic outliers in Europe.[1–4] On the other hand, the Saami, another Finno-Ugric speaking group, have been shown to be genetically distinct when compared with other European populations.[2–7]

The paucity of polymorphic markers has limited the use of the Y chromosome in population studies. However, several recently published microsatellites (eg[8,,9]), as well as some unique single base substitutions (eg[10,,11]), and a Y chromosomal Alu insertion (YAP) polymorphism[12,13] have increased the known Y chromosomal variation considerably. This has made Y chromosomal polymorphisms an efficient tool for supplementing and comparing the findings from maternally inherited mtDNA.

In a study on three Y chromosomal polymorphisms the Finns were found to be a strikingly homogeneous population.[14] On the other hand, the genetic diversity in the Swedes, Saami, and Estonians, who live in the geographical vicinity of the Finns, was higher. These results were interpreted as further evidence of a bottleneck in the expansion of the Finnish population around 2000 to 4000 years ago. The bottleneck or equivalent chance effects have been previously suggested by studies on recessive diseases existing with high frequency in the Finns, but absent or rare elsewhere.[15,16]

In the present study we further elucidate the genetic structure of the Finno-Ugric speaking populations by utilising a Y specific Alu insertion,[13] a recently described single base substitution polymorphism (Tat[11]), and six highly variable Y chromosomal microsatellite markers (eg[9]). The Y chromosomal variation in the Finno-Ugric speaking populations is compared with that found in Indo-European speakers and to some extent with that described in other populations used as a reference (Table 1; Figure 1).

Materials and Methods

DNA Samples

DNA samples were obtained from 502 unrelated male individuals (Table 1) from 11 Finno-Ugric speaking populations and five additional populations from Eurasia. Dutch, Inuit, and Pygmy reference populations were analysed previously.[17,18] Total DNA was purified from whole blood using a routine organic extraction method as described earlier.[19]

Analysis of the Y-specific Alu Insertion

The detection of the Y-specific Alu insertion was based on PCR amplification and subsequent agarose gel electrophoresis. The PCR was performed in a programmable heat block (MJ Research, Watertown, USA) as described earlier.[20] After amplification, the PCR-product was run on 2% agarose gel for 1 h at 90 V to detect the insertions. The results were recorded on Polaroid 667 ISO 3000/36° film under UV illumination after staining with ethidium bromide.

Analysis of the Tat Polymorphism

Screening for the Tat polymorphism was carried out by amplifying a 112 bp fragment using the primers Tat1 and Tat3,[11] and a subsequent digestion with NlaIII. The reaction mixture for the PCR consisted of 40 nmol of each dNTP, 10 pmol of each primer, 1.5 units of DynaZyme DNA polymerase (Finnzymes, Espoo, Finland) in 1 × DynaZyme buffer containing 10 mM Tris-HCl pH 8.8, 1.5 mM $MgCl_2$, 50 mM KCl, 0.1% Triton X-100. Each reaction consisted of 30 cycles with a time profile of 60 s at 95°, 60 s at 58°, and 60 s at 72°C. The detection of the restriction fragments was carried out on 2% agarose gels as described above. The C alleles were confirmed by amplification using the primers RD5 and RD3[11] and a NlaIII digestion. The PCR protocol was as described above for Tat1 and Tat3 primers, except that the annealing temperature was lowered to 53° and the elongation time was modified to 1 min 30 s.

Analysis of the Microsatellite Loci

Six Y chromosomal microsatellite loci were analysed. These were the GATA repeat DYS19,[8,21,22] the tetranucleotide

Table 1 The populations[30–34]

Population	*Linguistic group*	*Population size*	*Sample size*
Finns	Finno-Ugric	5 000 000	41
Saami	Finno-Ugic		
– Inari Saami		6 400	48
– Skolt Saami		500	27
Karelians	Finno-Ugric	70 000	56
Hungarians	Finno-Ugric	14 000 000	33
– Csangós		200 000	24
Mari	Finno-Ugric	550 000	39
Mordvins	Finno-Ugric		
– Mokshas		250 000	73
– Erzas		500 000	52
Ob-Ugrians	Finno-Ugric		
– Khanty		13 000	21
– Mansi		3 000	15
Yakuts	Altaic	300 000	12
Mongolians	Altaic	2 500 000	8
Koryaks	Chuckchi-Kamchatkan	8 000	10
Nivkh	not classified	4 500	9
Latvians	Indo-European	1 400 000	34
Dutch	Indo-European	13 500 000	88
Inuits	Eskimo-Aleut	47 000	62
Mbuti Pygmies	Nilo Saharan	30 000	31

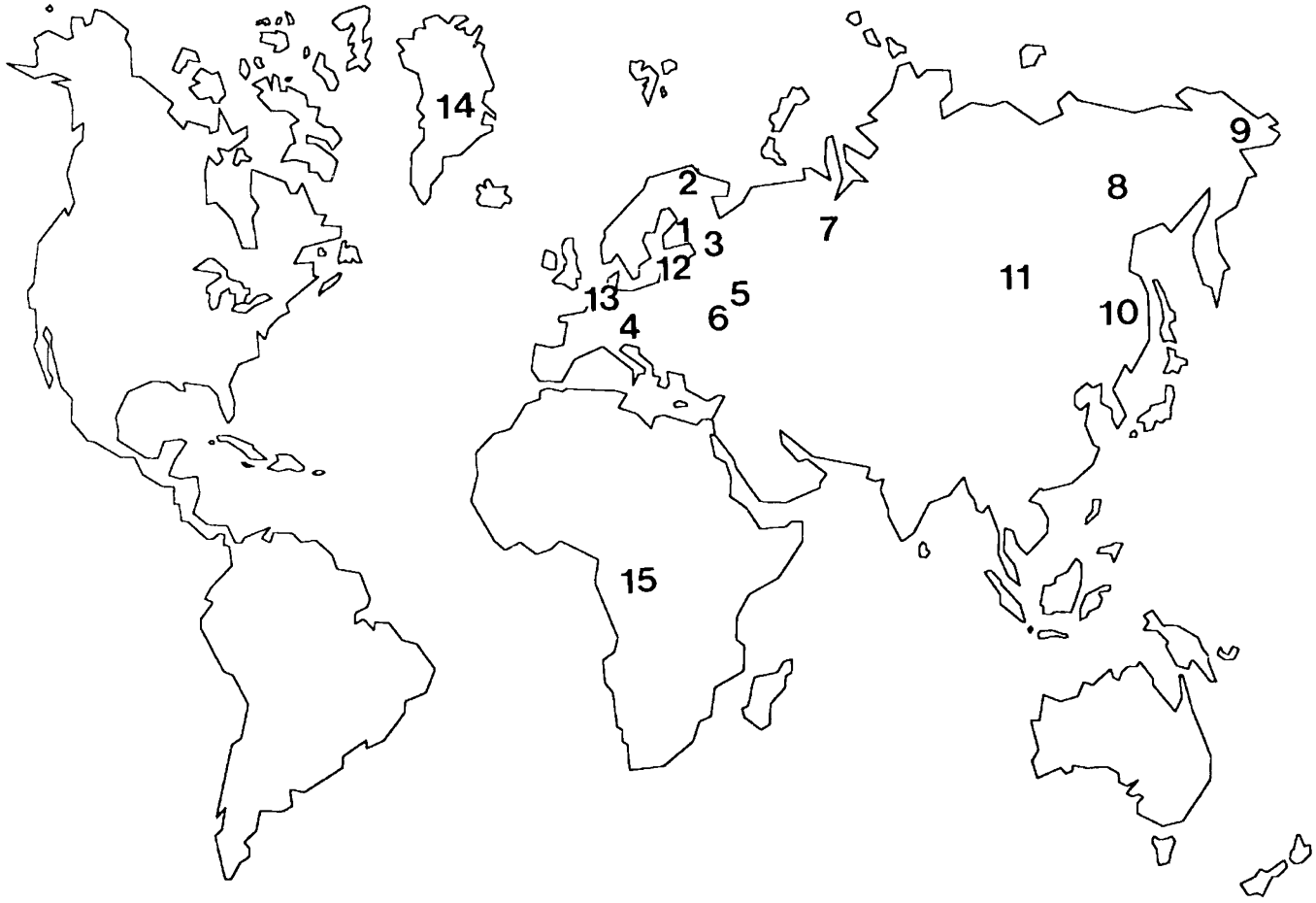

Figure 1 *The areas inhabited by the populations. 1: Finns, 2: Saami, 3: Karelians, 4: Csángós and other Hungarians, 5: Mari, 6: Mokshas and Erzas, 7: Mansi and Khanty, 8: Yakuts, 9: Koryaks, 10: Nivkh, 11: Mongolians, 12: Latvians, 13: Dutch, 14: Inuits, 15: Pygmies*

repeats DYS389-I and -II, DYS390, DYS393 and the trinucleotide repeat DYS392.[17] The markers were amplified by PCR as described earlier[8,17,22] using fluorescent primers applicable for the used sequencer. After amplification, the fragments were separated on 6% denaturing PAGE-gel on ALF Express (Pharmacia LKB, Uppsala, Sweden) or ABI Prism 377 (Perkin Elmer, Foster City, USA) automatic sequencers. The fragment lengths were analysed using Fragment Manager V3.0 Allelelinks (Pharmacia) or GeneScan (Perkin Elmer) software. In addition to internal and external molecular weight standards, we used locus-specific allele ladders with known sequences to determine unequivocally each genotype.

Networks Constructed from the Y Chromosomal Haplotypes

For each individual a Y chromosomal haplotype was constructed. A network of the observed haplotypes was constructed using information on the presence or absence of the YAP insertion, and the T and C alleles of the Tat polymorphism, as well as the alleles of each microsatellite marker. In a network only connections that could be explained by a single mutation step difference were taken into account. As a single step we allowed an increase or decrease of a single repeat at one of the microsatellites.

Statistical Analysis

The genetic distances derived from the information on the haplotypes consisting of the six microsatellite markers were calculated using the Arlequin 1.1 program package (http://anthropologie.unige.ch/arlequin). Of the two possible molecular distance methods, number of different alleles and sum of squared size differences, the latter was selected as it incorporates more molecular information. The distances were compared both by subdividing (where applicable) the samples according to the Tat polymorphism in C or T frames and without this subdivision. They were further linearised with respect to time using the transformation $D = -\ln(1 - F_{ST})$.[23] The variances, gene and haplotype frequencies, and molecular genetic diversity values and their *P* values were also computed using the Arlequin 1.1 program. Neighbour-joining trees were constructed from the linearised genetic distance matrices with program Neighbor of the Phylip 3.572c package.[24] The haplotype F_{ST}-(R_{ST}) and their *P* values (derived from approximately 1000 permutations) were obtained using Amova implemented in the Arlequin program package.

Estimates on the Tat C-mutation age were made using program BDMC vers. 2.1[25] including the north-Eurasian population data. The mutation rate for microsatellite markers was set to 0.002 as reported for tetra and trinucleotide repeats.[26] An overall estimate maximising the likelihood over growth and selection coefficient as well as over the generation time with 5000 Monte Carlo replicates and generation interval varying from one to five generations was computed.

Results

Yap

Our Hungarian sample consisted of a set of individuals of mixed Hungarian origin from Budapest and a separate sample of a single Hungarian ethnic group, the Csángós. A high occurrence of the YAP insertion was found in both Hungarian groups (37.5% in the Csángós and 7.5% in the Hungarians from the Budapest area; Table 2). The insertion was present in low frequency in the Karelians (5%), the Latvians (3%), and the Dutch (2.3%), but it was not detected in the other Eurasian populations, including the Finns and the Saami.

Table 2 Frequencies of Yap$^+$ and Tat C alleles

Population (n)	*Yap*$^+$	*Tat C*
Finns (41)	0.00	0.611
Saami		
– Inari (48)	0.00	0.417
– Skolt (27)	0.00	0.617
Karelians (56)	0.05	0.396
Hungarians		
– Budapest (33)	0.18	0
– Csángós (24)	0.38	0
Mari (39)	0	0.333
Mordvins		
– Mokshas (73)	0	0.082
– Erzas (52)	0	0.120
Ob-Ugrians		
– Khantys (21)	0	0.632
– Mansi (15)	0	0.182
Yakuts (12)	0	1.0
Mongolians (8)	0	0
Koryaks (10)	0	0.200
Nivkhs (9)	0	0
Latvians (34)	0.03	0.294
Dutch (88)	0.02	nd
Inuits (62)	0	nd
Pygmies	0.37	nd
Italians[12,13/11]	0.11	0
Africans[13,34/11]	0.74	0
Japanese[13/11]	0.24	0
Chinese[13/11]	0.00	0

Tat Polymorphism and Microsatellite Variation

The C allele of the Tat polymorphism was present in most of the Finno-Ugric speaking populations. The frequency of the C allele ranged from 63.2% in the Khanty to 8.2% in the Mokshas, whereas the allele was absent in the Hungarians (Table 2). It was also common in the Indo-European speaking Latvians (29.4%) and monomorphic in the Yakuts.

Complete haplotypes that consisted of the YAP element, the Tat polymorphism, and all six microsatellite loci could be constructed for 450 samples (626 including the Dutch, Inuit, and Pygmy reference samples). A total of 222 different haplotypes were found (Appendix: obtainable on request by E-mail from the corresponding author). Of these haplotypes, 169 were unique to a single population, 45 were found in at least two populations and 20 of these in at least three populations. The most widely distributed haplotype was found in eight populations. This haplotype carried the T allele of the Tat polymorphism.

Four loosely formed groups could be seen in a network constructed from the haplotypes (data not shown). One of the groups was formed of the Tat C allele haplotypes, whilst the others carried the T allele. Fifty-four of the 222 haplotypes could not be connected to any of these groups by a single mutational step.

The polymorphism found in the Tat enabled us to subdivide the Finno-Ugric populations and Latvians according to the presence of T or C mutation (see available Appendix). The average gene diversities computed within the subdivisions were suggestively different, the C mutation group showing less diversity (value range for T = 0.315–0.634 and for C = 0.100–0.400 with overall diversities for the T group of 0.588 ± 0.332 and for the C group of 0.374 ± 0.229; Table 3) in all population groups where the polymorphism was present. This strongly suggests that the C mutation is more recent than the T mutation, which is thus likely to be the ancestral one.

Without the subdivision into the Tat groups C and T the variation within the populations covered 75.3% of the total variation, whilst the subdivision into these groups considerably lowered the within variation to 47.9%. The general diversity values calculated from the number of different alleles between the haplotypes in the undivided populations were not significantly different among the Finno-Ugric speaking populations. They were highest in the Karelians (0.596) and lowest in the Mansi (0.384).

Table 3 Genetic diversities with standard deviations

Population	*Diversity*	*SD*	*n*
Finns	0.489	0.291	39
Finns C	0.225	0.161	24
Finns T	0.448	0.283	15
Inari Saami	0.455	0.273	47
Saami C	0.176	0.134	29
Saami T	0.315	0.211	18
Skolts	0.544	0.324	24
Skolts C	0.152	0.129	10
Skolts T	0.353	0.235	14
Karelians	0.596	0.341	53
Karelians C	0.279	0.190	21
Karelians T	0.478	0.287	32
Mari	0.569	0.331	37
Mari C	0.282	0.198	13
Mari T	0.534	0.319	24
Mokshas	0.408	0.250	49
Mokshas C	0.111	0.125	4
Mokshas T	0.349	0.221	45
Erzas	0.507	0.298	50
Erzas C	*0.400*	0.291	6
Erzas T	0.445	0.268	44
Ob-Ugrians	0.447	0.271	35
Ob-Ugrians C	0.192	0.148	14
Ob-Ugrians T	0.440	0.273	21
Khanty	0.461	0.285	20
Mansi	0.384	0.250	15
Hungarians (T)	0.582	0.340	30
Csángós (T)	0.508	0.306	23
Latvians	0.556	0.325	34
Latvians C	*0.100*	0.098	10
Latvians T	0.491	0.297	24
Yakuts (C)	0.152	0.127	12
Koryaks	0.648	0.402	10
Mongols	0.506	0.335	8
Nivkhs	0.634	0.399	9
Dutch	0.539	0.311	88
Inuits	0.486	0.287	62
Pygmies	0.604	0.350	31

Haplotype Diversity in Finns and Saami

The Finns represented 15 haplotypes. One of them was found in nine (25%) individuals. This haplotype carries the Tat C allele and is identical with the haplotypes 3 and 17 described earlier[11] at the six loci tested. Altogether 20 individuals (55.5%) could be classified into this haplotype, or haplotypes that were separated from it by one mutational step only (Figure 2). The majority (5/9) of the remaining haplotypes, making 33.3% of the total, were clustered around a haplotype harbouring the Tat T allele. All of the three most common haplotypes in the Finns were found in at least three other Finno-Ugric speaking populations, two of them also in single Indo-European individuals.

The Saami are classified into two groups. In 47 Inari Saami samples, 13 haplotypes were observed. The most common haplotype (Tat C) was found in 13 individuals (27.7%). Ten other individuals harboured haplotypes that differed from this haplotype by one mutational step only. Thus, altogether 23 individuals (47%) of the Inari Saami represented a single common lineage. The most common Finnish haplotype was also frequent in the Inari Saami (10.2%). It was restricted to the Finns, the Saami, and the Karelians, with the exception of one Koryak.

The Skolt Saami also showed two major lineages. Two haplotypes, one harbouring the Tat C allele and the other Tat T, were both present in eight out of 23 samples (34.8%). The rest of the haplotypes were found only once or twice. Thus, 16/23 (69.6%) of the Skolt Saami had a background of these two lineages, which are also found in the Saami from the Inari Lake region. The former of the haplotypes was also observed in two Karelian samples.

Haplotype Diversity in Other Populations

In the Karelians and Latvians, who live to the east and south, respectively, in the geographical vicinity of Finland, a large number of haplotypes were found. In both populations the haplotypes were divided by the presence of the C allele of the Tat polymorphism (Table 2). The most common Karelian haplotype was shared with the Finns and the Saami. The C allele was present in 10/34 (29.4%) of the Latvian samples. Both Karelian and Latvian samples were characterised by high diversity of haplotypes, with a large number of unique haplotypes. Furthermore, no such grouping of haplotypes as was seen in the Finns and the Saami (Figure 2) was observed.

None of the Hungarian haplotypes carried the Tat C allele. Both Hungarian groups showed a high frequency of the YAP insertion. The groups shared three haplotypes out of 29. The haplotype diversity was higher in the Budapest area than among the Csángós.

The haplotype diversity was reduced in other Finno-Ugric populations with the exception of the Mari. The C allele was clearly most frequent in the Ob-Ugrian Khanty (63.2%). It was also common in the Mari (33.3%).

The Yakut did not share any haplotypes with any other of the populations. The Tat C allele was monomorphic in our sample and is common in the Yakuts, as seen in earlier studies.[11] The diversity of Yakut haplotypes was the lowest (0.152, Table 3) when undivided populations were considered, but was within the normal range for C haplotypes. Among Siberian populations the C allele was found also in the Koryaks, but not in the Nivkh.

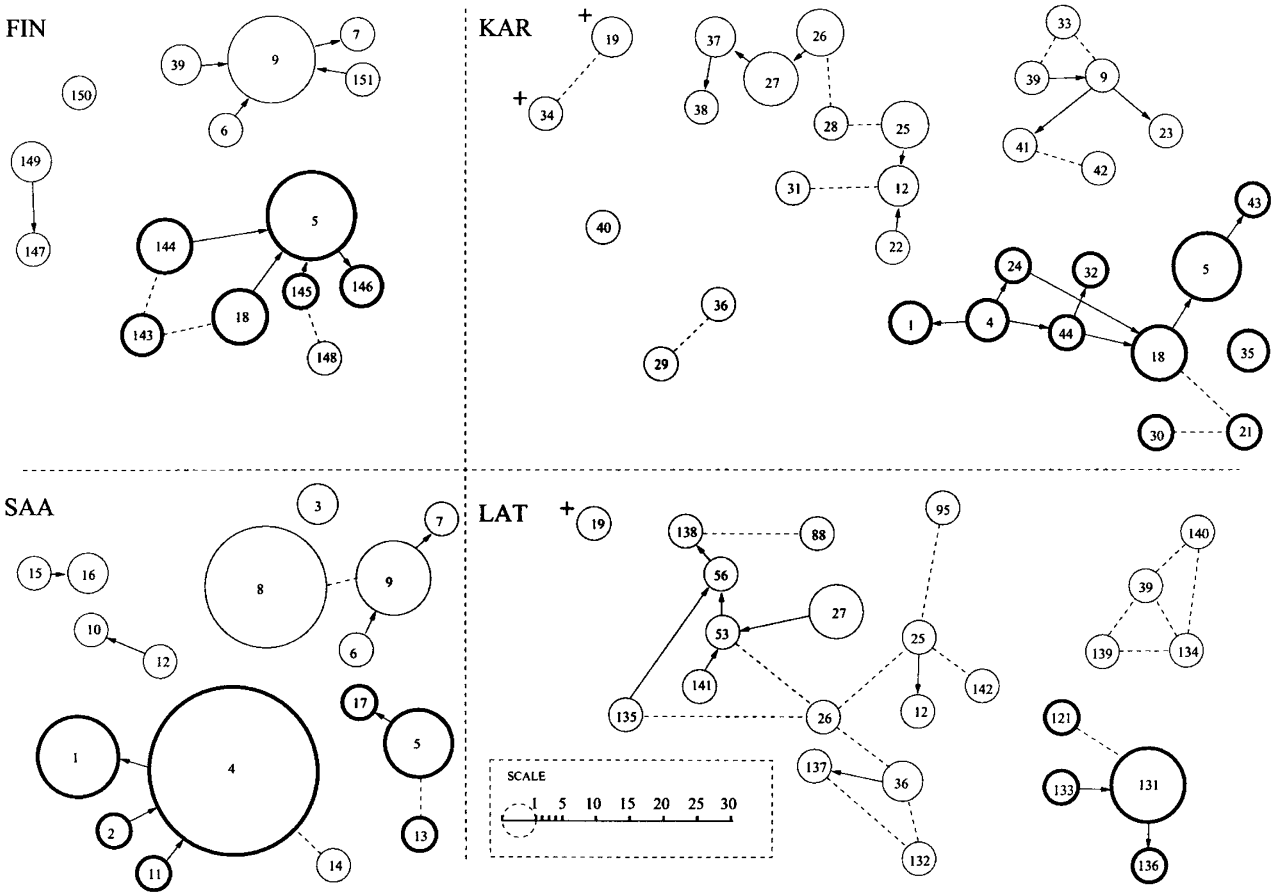

Figure 2 *Haplotypes found in the Finns, Saami, Karelians, and Latvians. Bold circles indicate the haplotypes harbouring the C allele of the Tat polymorphism. Connections of single repeat mutational differences are shown between the haplotypes by arrows indicating the direction of the allele size growth. The dotted lines illustrate connections of two different alleles. The presence of the Alu insertion is marked by +. (The numbers refer to identification numbers of the haplotypes in the available Appendix.)*

The Dutch belonged to the heterogeneous populations in this study. Among the Dutch, a majority of the individuals formed a network of related haplotypes.

Age of Tat C

The age estimate for Tat C including all the northern Eurasian typed populations gave a maximum likelihood result indicating that its age is relatively young, 222 generations (Figure 3). Given a generation time of 20 years this would translate to 4440 years since its introduction into the populations (approximate 95% confidence interval 157–310 generations or 3140–6200 years). The maximum estimate was obtained with a growth-selection coefficient 0.055, which seems reasonable since many of the populations can be considered relatively static with regard to expansion.

Genetic Distances

Only those non-Finno-Ugric populations where the sample size was considered reasonable ($n \geq 15$) were included in the genetic distance computations (Table 4). The Ob-Ugrian Khanty and Mansi who showed non-significant distance (0.06) were combined into one Ob-Ugrian group before division into C and T haplotype groups. The pairwise genetic distance values in the undivided populations indicated a very close relationship between the Finns and the Saami. However, the subdivision into Tat C and T haplotypes showed that the closeness is mostly due to T haplotypes. This situation was reversed for the Finns and Karelians where the C groups were close to each other, whilst the T haplotypes showed more differences. The Finnish C haplotype group did not show any closer relationship to the Finnish T-group than to the other populations'

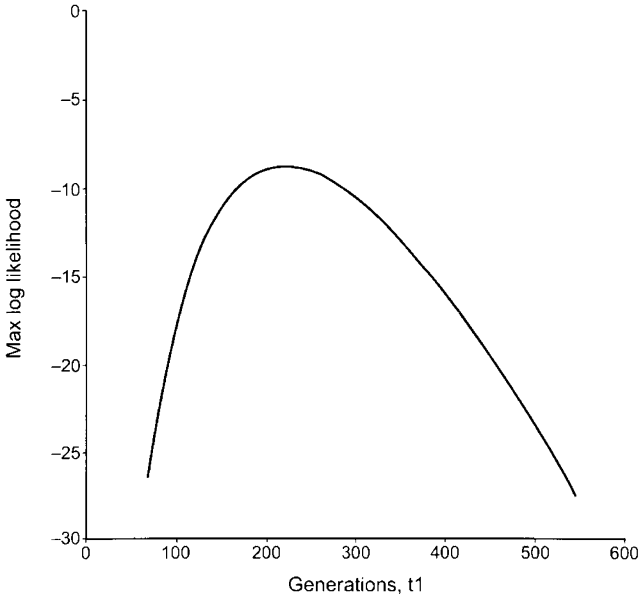

Figure 3 *Age estimation of Tat C mutation using program BDMC version 2.1.[25] The corresponding age in number of generations resulting in the maximum likelihood point estimate shown on graph y-axis was obtained using different values for the combined population growth and selection coefficient (xi, not shown). The number of Monte Carlo replicates for a given maximum likelihood point estimate was 5000 and was computed at several xi intervals from 0.01 to 0.2 to refine the estimate while keeping all other parameters constant. The program option to evaluate conditional on the observed number of Tat C mutations was used. The overall likelihood is maximised at xi = 0.055 with generation number equalling 222. Approximate 95% confidence interval (2 log units) was 157–310 generations*

T-groups with the exception of the Ob-Ugrians whose relationship to C haplotype groups were generally closer. The same applied to all other C/T subdivisions within the populations. Among the C groups the Latvians were most distinct from the others.

In the phylogenetic trees based on R_{ST} distances of undivided populations (Figure 4, inset) there was a poorly defined structure. However, when the Tat C/T sub-grouping was applied (Figure 4, main tree) the tree showed that all the Finno-Ugrians within the C subdivision tended to group closely to each other whilst the T subdivision appeared to have a more complex structure. Here also the Skolts, Erzas, Finns, and Inari Saami comprised a close grouping, whilst the Mari, Mokshas, and Karelians showed a closer relationship to the Hungarians and Latvians. Within the C group of haplotypes the Latvians had the most distant relationship, although still showing a tendency to belong to the other C groups. The Inuits also grouped with the C haplotype sub-populations as did the Ob-Ugrian T haplotypes which appeared at the branch leading to the C haplotype groups.

Discussion

Two Major Founding Male Lineages in Finns and Saami

Genetic studies of recessive diseases found in the Finns, but rare or absent elsewhere have demonstrated the effect of local isolation on the gene pool of a small founding population.[14,15,27] Recently, studies on non-functional parts of the mitochondrial and Y chromosomal genomes have also suggested a bottleneck in the Finns.[13]

To supplement our previous studies on maternal mtDNA lineages[4,6] we have analysed the population structure of the Finns and other Finno-Ugric speaking populations using Y chromosomal microsatellite markers and a Y specific Alu insertion. A recently described single base substitution (Tat polymorphism) was also analysed in order to anchor the hypervariable microsatellites to slowly evolving sites in the Y chromosome.

The C allele at the Tat polymorphic site was recently found to be frequent in the Finnish population.[11] Our analysis further shows that the Tat polymorphism divides the Finns into two major lineages, both were showing a low level of variation. The first lineage (60% of the samples) consisted of haplotypes carrying the C allele. Furthermore, 95.5% of these haplotypes were either identical or differed by a single mutational step at one microsatellite locus. The second founding lineage was comprised of haplotypes carrying the T allele. Again, a cluster was formed around a single haplotype, including 25% of the total number of Finnish haplotypes. Thus, our data point to two major founding lineages, which have survived to the present-day Finnish population. This pattern of haplotype diversity is clearly different from those found in, for example, the Dutch, Hungarians, and even Karelians, where the Y chromosomal lineages were represented by distribution into several haplotypes, and where in most cases a given haplotype is seen in a single individual only.

A similar picture is found when the samples from the two Saami populations are analysed. The Inari Lake sample consists of individuals who live in the same area in Northern Finland, but belong to originally different Saami groups in Finland, mainly the Lake and Mountain Saami. The Lake Saami subsist from fishing, whereas the Mountain Saami have traditionally been

Table 4 R_{ST} distances (× 100) between the undivided populations with $n \geq 15$ (upper triangle), and those divided according to Tat C/T mutation (lower triangle). Distances, which were non-significant (based on about 1000 permutations) at 5% level are underlined. (The population symbols are as shown in the available Appendix.) Tat C/T appears as c,t

FINc							FIN	INA	SKO	KAR	MAR	MOK	ERZ	XAN	MAN	LAT	CSA	HUN	DUT	INU	PYG	
INAc	29							2	12	5	16	50	33	9	5	19	27	28	11	17	42	FIN
SKOc	35	1							7	10	17	54	27	16	8	22	29	32	10	28	39	INA
KARc	6	16	16							9	5	43	7	17	9	13	13	18	5	36	17	SKO
MARc	45	35	22	32							5	27	22	14	10	5	9	9	6	21	29	KAR
MOKc	53	50	54	36	0							32	13	18	10	11	4	8	6	29	25	MAR
ERZc	38	42	35	16	3	18							37	60	60	16	28	9	39	62	34	MOK
OBIc	36	42	32	14	24	31	1							39	32	19	15	18	17	52	7	ERZ
LATc	66	59	57	55	30	77	53	57							6	29	38	38	27	21	45	XAN
FINt	82	81	76	79	63	70	72	77	76							25	29	32	12	23	42	MAN
INAt	84	84	79	82	70	76	77	80	79	3							12	5	14	42	20	LAT
SKOt	77	77	66	73	56	56	58	68	63	9	11							2	9	41	22	CSA
KARt	66	67	60	65	55	55	56	64	60	17	30	23							15	43	22	HUN
MARt	71	71	63	68	54	56	55	64	63	19	32	14	6							32	32	DUT
MOKt	86	87	86	86	81	85	83	86	85	61	68	57	17	38							60	INU
ERZt	71	69	61	68	56	58	59	65	58	9	9	0	24	18	51							
OBIt	44	40	28	37	12	12	15	32	28	34	45	32	29	27	61	36						
LATt	71	73	66	70	60	61	61	69	64	29	40	30	1	17	7	31	35					
CSAt	67	68	59	64	49	50	51	60	60	14	29	17	6	2	42	23	19	16				
HUNt	60	62	54	59	47	46	47	57	53	17	30	21	0	4	19	24	21	2	2			
DUT	47	42	37	44	32	30	36	44	40	12	24	25	19	20	48	26	8	28	9	15		
INU	29	35	33	28	34	26	25	28	51	57	63	60	49	48	70	60	27	56	41	43	32	
PYG	66	65	55	63	52	52	51	59	50	18	16	3	22	19	40	7	34	22	22	22	32	60
	FINc	INAc	SKOc	KARc	MARc	MOKc	ERZc	OBIc	LATc	FINt	INAt	SKOt	KATt	MARt	MOKt	ERZt	OBIt	LATt	CSAt	HUNt	DUT	INU

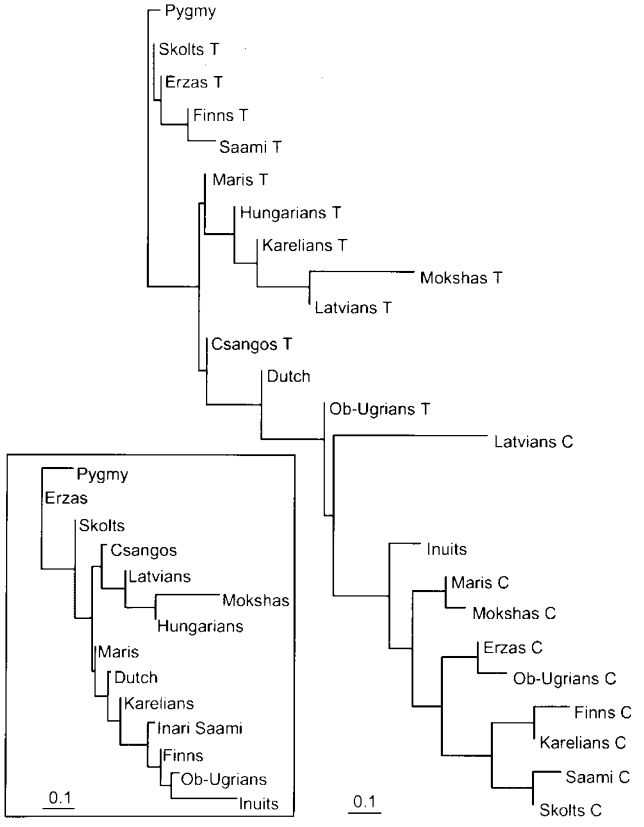

Figure 4 *A neighbour-joining tree constructed from the genetic distances with TAT C/T subdivided population data shown in Table 4. Inset shows a tree obtained without the subdivision of the data. Pygmies were selected as an outgroup for the trees*

reindeer-herding nomads. Furthermore, both groups speak a different Saami language. The Skolt Saami are a uniform group identified by their own language and Greek Orthodox religion. They have lived in comparative isolation in the forested areas east of the Inari Lake on both sides of the present Finnish–Russian border, with fishing as their main livelihood. Their numbers have always been small, rising to approximately 500 individuals.[28]

In the Inari Saami a single haplotype carrying the C allele and haplotypes related to it by a single repeat mutational step were present in approximately half the samples (48.9%). The second main group (10.6%) in the Inari sample was built around one haplotype that harbours the C allele. This group is also found in the Finns. Similarly, the Skolt Saami had two lineages that can be considered as founding lineages. Four of the five most common lineages in the Saami are shared with the Karelians and two with the Finns. Considering the close geographical vicinity of these populations, the sharing of Y chromosomal lineages is not surprising. A low level of admixture between these populations has also been seen in the mtDNA lineages[4,6] and protein polymorphisms in these populations.[29] The two different Saami groups share their most frequent haplotypes, indicating common ancestry or admixture. Our data also suggest the presence of greater admixture between the Inari Lake Saami and the Finns than between the Skolt Saami and the Finns. The Finns and the Saami may have gained the C allele of the Tat polymorphism from a common source or it may just be a result of later admixture.

The finding of only a few Y chromosomal lineages in the Finns and the Saami is in accordance with the general view of the population history of Finland. This large country was probably populated by only a small number of individuals for relatively long periods, and the population may also have reduced in size at times during its history. The settlement history of Finland, with most of the people living in small groups relatively separate from each other until only a few decades ago ('isolation by density'[27]) may certainly have been the driving force of the founding effects and genetic drift. The same is true of the Saami, whose historical documents show that the population has always been small, and has experienced expansion only in recent decades.[28]

The Karelians are close neighbours of the Finns and speak a language which some consider a dialect of Finnish. The Karelian haplotypes were heterogeneous compared with the Finns (Figure 2), possibly due to greater Slavonic admixture in the Karelians and their closer geographical relationship to Slavonic peoples; there is a more pronounced founding effect in the Finns.

Other Populations

The Y specific Alu insertion, located in the long arm of the Y chromosome, is a member of the Alu family of repeated DNA elements. The insertion is predominant in sub-Saharan African populations, whereas it is absent in most north Asian populations, with the notable exception of the Japanese. In Europe it is found in approximately 10% of Y chromosomes.[13,20] In the present study, the Y specific Alu insertion was unexpectedly common in the Hungarians (17.5%), especially in the Csángó group (37.5%). The Alu insertion haplotypes found among the Karelians and the Latvians were not present in the Hungarians, although haplotypes with one mismatch only were found. Therefore, it seems likely that part of the Yap insertion

haplotypes found among the Hungarians have originated from the same source as the insertion haplotypes in Northern Europe.

The diversity among the Csángós was somewhat smaller than among other Hungarians. The Csángós, as well as the other Hungarians lacked the C allele of the Tat polymorphism, which is present in many of the northern Eurasian populations. The linguistic relationship between the Finns or Saami and the Hungarians is not discernible at the genetic level, possibly due to later dispersal or founder effects.

Haplotype clustering and decrease in haplotype diversity were seen in some reference populations, most notably in the small sample of Yakuts, whose haplotypes are at the same time less diverse and strikingly different from the other populations. Other populations in which were found clustering of haplotypes, decrease in haplotype diversity, and therefore possible founder effect, are the Inuit and the Pygmies.

Origin of the Tat C Allele?

It has been suggested[11] that the C allele had its origin in the area around present-day Northern Mongolia, and that the presence of the allele in the Finns would indicate substantial Asian influences in the Finnish males. This suggestion was mainly based on the observation that, in the populations with highest C allele frequency, the haplotypes carrying the allele varied more in Mongols than in Finns. The C-allele was also found in Estonians and Mari. Our data demonstrate that the C allele is present in all the Finno-Ugric speaking populations studied so far, with the exception of the Hungarians. Furthermore, the highest frequency of the C allele (63.2%), with the exception of the monomorphic Yakuts, was found in the Ob-Ugrian Khanty, living on the western slopes of the Ural Mountains. In addition, the C allele is present in Baltic and Scandinavian populations.[11] Furthermore, haplotypes harbouring the C allele in other populations are present in 69.2% of the Inuits. The mutation was common (29.4%) in the Latvians, an Indo-European speaking population. On the other hand, it has not been found, for example, in the German, British, Italian or Basque populations.[11] In this study, the only microsatellite haplotype with both T and C alleles, was haplotype 11 found in the Ob-Ugrians (T allele), Saami (C allele), Mokshas (C allele), and the Inuit (allelic status unknown). If this is not a result of convergent evolution of microsatellite alleles, the haplotype 11 may present the original C haplotype, or one close to it. The present data do not allow us to point to a specific geographical site as the origin of the mutation, nor the direction of the gene flow. The wide distribution of the mutation in the north Eurasian populations suggests that it arose in Northern Eurasia and subsequently spread to various populations. Its enrichment in certain geographically distant populations can be explained by phenomena such as bottleneck and genetic drift.

The age estimate of Tat C presented here is the first we know of and may contain inaccuracies due to selection of populations, as it was restricted to those populations we had available for Tat typing. It is obvious that this mutation must have been introduced from an unknown population where it originated into all the others by admixture. Furthermore, we do not know the original haplotype in which the mutation arose. It was therefore not possible to estimate the actual age in the individual populations in which it is present. We therefore considered our sample of Tat-typed populations together as a sample of a northern Eurasian super-population where this mutation supposedly arose and which may be assumed to contain most of the haplotype variation that has emerged since the mutation event. Absence of some of the actual variation is bound to lead to an underestimate of the age. To date, the age estimate, 222 generations or 4440 years, has relatively wide confidence limits (3140–6200 years). This can be narrowed when more data accrues. Judging from the geographic distribution of the mutation, the upper age boundary would be expected to be no earlier than the end of the last glacial period around 12 000 years ago. The present estimate would therefore fit well within the time frame posed by this major climatic event.

Conclusion

Our data suggest the reduced variation arises from a bottleneck in the male lineages of the Finns and the Saami or drift due to a small effective population of males over a longer period. Furthermore, it reveals two major founding lineages in both present populations. One of the lineages is characterised by a C allele in the recently described Tat polymorphism. This allele was recently suggested to have originated in Central Asia, thus indicating a substantial Asiatic influence in Finnish males. In contrast, our data suggest that the C allele is common in many of the north Eurasian populations. Given the relatively fast evolving microsatellite markers it would seem that this mutation is a relatively recent one. It is conceivable that the high frequency in

the Finns may have emerged by way of a founding effect with its source in contacts with the Baltic populations, an assumption which would be in accordance with the archaeological and some recent blood group data. However, this hypothesis is not supported by the higher microsatellite diversity of the Finnish C group, which would indicate that the introduction of the mutation into the Finns predates that into the Balts.

Acknowledgements

We wish to express our thanks to all our blood donors. Tuula Koski of Lappeenranta, Finland, Michael Voevoda of the Russian Academy of Science, Novosibirsk, and Andrei Kozlov of the ArctAn-C Innovative Laboratory gave us invaluable help in arranging sample collection in Russia. Tiiu Anttinen from National Bureau of Investigation, Finland, Galina Verschubskaja from ArctAn-C Innovative Laboratory, and Dr Vladimir G Durnyagin, Deputy Chief, Surround Traveling Clinic, Khanty-Mansisk, Russia also helped us in the sample collection. Professor Astrida Krumina from the Medical Academy of Latvia provided us with the Latvian samples. Professor LL Cavalli-Sforza and LF Bernini kindly provided the Pygmy samples. Ilona Carlsson and Pirkko Jalava gave us excellent technical assistance in sample analysis at the department of Medical Genetics in Turku. Dr Jukka Heikkonen from the Helsinki University of Technology wrote numerous programs for us to make data handling easier. Dr Bruce Rannala kindly informed us of the new version 2.1 of the program BDMC, correcting some bugs of the earlier versions. This work was supported by grants 854479 and 859236 from the Academy of Finland and a grant from the Maud Kuistila Foundation.

References

1 Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S: Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European Populations. *Biol Chem Hoppe-Seyler* 1994; **375**: 837–840.
2 Sajantila A, Pääbo S: Language replacement in Scandinavia. *Nat Genet* 1995; **11**: 359–360.
3 Sajantila A, Lahermo P, Anttinen T *et al*: Genes and languages in Europe: an analysis of mitochondrial lineages. *Genome Res* 1995; **5**: 42–52.
4 Lahermo P, Sajantila A, Sistonen P *et al*: The genetic relationship between the Finns and the Finnish Saami (Lapps): Analysis of nuclear DNA and mtDNA. *Am J Hum Genet* 1996; **58**: 1309–1322.
5 Cavalli-Sforza LL, Piazza A: Human genomic diversity in Europe. A summary of recent research and prospects for the future. *Eur J Hum Genet* 1993; **1**: 3–18.
6 Nei M, Roychoudhury AK: Evolutionary relationships of human populations on a global scale. *Mol Biol Evol* 1993; **10**: 927–943.
7 Cavalli-Sforza LL, Menozzi P, Piazza A: *The History and Geography of Human Genes*. Princeton University Press: Princeton, 1994.
8 Roewer L, Arnemann J, Spurr NK, Grzeschik KH, Epplen JT: Simple repeat sequences on the human Y chromosome are equally polymorphic as their autosomal counterparts. *Hum Genet* 1992; **89**: 389–394.
9 Jobling MA, Tyler-Smith C: Fathers and sons: the Y chromosome and human evolution. *TIG* 1995; **11**: 449–456.
10 Underhill PA, Jin L, Zemans R, Oefner PJ, Cavalli-Sforza LL: A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc Natl Acad Sci USA* 1996; **93**: 196–200.
11 Zerjal T, Dashnyyam B, Pandya A *et al*: Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 1997; **60**: 1174–1183.
12 Persichetti F, Blasi P, Hammer M *et al*: Disequilibrium of multiple markers on the human Y-chromosome. *Ann Hum Genet* 1992; **56**: 303–310.
13 Hammer MF: A recent insertion of an Alu element on the Y-chromosome is a useful marker for human population studies. *Mol Biol Evol* 1994; **11**: 749–761.
14 Sajantila A, Salem A-H, Savolainen P, Bauer K, Gierig C, Pääbo S: Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. *Proc Natl Acad Sci USA* 1996; **93**: 12035–12039.
15 de la Chapelle A: Disease mapping in isolated human populations: the example of Finland. *J Med Genet* 1993; **30**: 857–856.
16 Peltonen L, Pekkarinen P, Aaltonen J: Messages from an isolate: Lessons from the Finnish gene pool. *Biol Chem Hoppe-Seyler* 1995; **376**: 697–704.
17 Roewer L, Kayser M, Dieltjes P *et al*: Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum Mol Genet* 1996; **5**: 1029–1033.
18 de Knijff P, Kayser M, Caglia A *et al*: Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 1997; **110**: 134–149.
19 Vilkki J, Savontaus M-L, Nikoskelainen EK: Human mitochondrial DNA types in Finland. *Hum Genet* 1988; **80**: 317–321.
20 Hammer MF: A recent common ancestry for human Y chromosomes. *Nature* 1995; **378**: 376–378.
21 Arnemann J, Jakubiczka S, Schmidtke J, Schäfer R, Epplen JT: Clustered GATA repeats (Bkm sequences) on the human Y chromosome. *Hum Genet* 1986; **73**: 301–303.
22 Santos FR, Pena SDJ, Epplen JT: Genetic and population study of a Y-linked tetranucleotide repeat DNA polymorphism with a simple non-isotopic technique. *Hum Genet* 1993; **90**: 655–656.
23 Reynolds J, Weir BS, Cockerham CC: Estimation for the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983; **105**: 767–779.
24 Felsenstein J: PHYLIP (Phylogeny inference package), version 3.57c. University of Washington: Seattle, 1995.
25 Slatkin M, Rannala B: Estimating the age of alleles by use of intraallelic variability. *Am J Hum Genet* 1997; **60**: 447–459.

26 Weber JL, Wong C: Mutation of human short tandem repeats. *Hum Mol Genet* 1993; **2**: 1123–1128.
27 Nevanlinna HR: The Finnish population structure: a genetic and genealogical study. *Hereditas* 1972; **71**: 195–236.
28 Aikio S, Aikio-Puoskari U, Helander J: *The Saami Culture in Finland*. Publications of Lapin sivistysseura 49. Valtion painatuskeskus Oy, Helsinki, 1994.
29 Guglielmino CR, Piazza A, Menozzi P, Cavalli-Sforza LL: Uralic genes in Europe. *Am J Phys Anthrop* 1990; **83**: 57–68.
30 Ruhlen M: *A Guide to the World's Languages*. Stanford University Press: Stanford, California, 1987.
31 Lehtivirta J, Seurujärvi-Kari I: Saamelaiseet. In: Laakso J (ed.). *Uralilaiset kansat: tietoa suomen sukukielistä ja niiden puhujista*. WSOY: Helsinki, 1991, pp 123–155.
32 Grimes BF (ed.): *Ethnologue Languages of the World*. www.sil.org/ethnologue/, 1996.
33 Jahunen J, Salminen T: *UNESCO Red Book Report on Endangered Languages. North-East Asia*. http://www.helsinki.fi/tasalmin/nasia_index.html, 1993–1996.
34 Spurdle A, Jenkins T: The Y Alu polymorphism in southern African populations and its relationship to other Y-specific polymorphisms. *Am J Hum Genet* 1994; **54**: 319–330.