

Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages

M. Hedman^a, V. Pimenoff^a, M. Lukka^b, P. Sistonen^c, A. Sajantila^{a,*}

^aDepartment of Forensic Medicine, Laboratory of Forensic Biology, University of Helsinki, P.O. Box 40, 00014 Helsinki, Finland

^bNational Public Health Institute, Mannerheimintie 166, 00300 Helsinki, Finland

^cFinnish Red Cross Blood Transfusion Service, Kivihaantie 7, 00310 Helsinki, Finland

Received 6 April 2003; received in revised form 29 July 2003; accepted 31 July 2003

Abstract

We analysed samples of 400 Finnish males using nine Y-chromosomal short tandem repeat (STR) loci (minimal haplotype); for 200 of these subjects an additional seven Y-chromosomal STR loci were used. The geographical distribution of the observed haplotypes was determined from 200 individuals of known paternal origin within Finland. The observed number of alleles varied from 2 to 13 alleles per locus. A total of 146 minimal haplotypes were identified in our population sample. Interestingly, 90 (22.5%) individuals shared an identical haplotype. This haplotype was extremely frequent in the northern and eastern subpopulations of Savo, Pohjanmaa and Karjala (53, 42 and 37%, respectively). With the seven additional loci analysed in the sample of 200 individuals, 120 haplotypes were identified, and individuals sharing the most common haplotype decreased to 13.0%. However, in comparison to other European populations, the Finnish population showed decreased genetic diversity (GD) when the number of different minimal haplotypes in the population was divided by the sample size (36.5% in Finns versus 83.7% on average). Our results strongly support the earlier hypothesis of individual isolated Y-chromosomal lineages and population substructuring in Finland. For paternity testing, power of exclusion was 92% using minimal haplotype data, but including the seven additional loci this value increased to 97%.

© 2003 Elsevier Ireland Ltd. All rights reserved.

Keywords: Y STR; Forensic DNA; Finnish population; Population genetics

1. Introduction

The Y chromosome is inherited from father to son and the majority of its DNA lacks the potential for recombination, barring the pseudoautosomal regions in the telomeric parts of the chromosome [1]. Therefore, all descendants of a male lineage share a common Y-chromosomal haplotype in the non-recombining part, unless a mutation alters the haplotype. The average mutation rate for Y short tandem repeat (STR) loci was recently estimated to be 3.17×10^{-3} , although marked variation was present in the locus-specific mutation rate, which ranged from 0 to 8.58×10^{-3} [2]. Due to a relatively high mutation rate, the Y-chromosomal STRs

provide a resolution that is useful as a forensic tool for solving specific problems, such as identification of individuals or paternity testing, in cases where the samples in question putatively share a male lineage. The online population reference database of European, North American and Asian Y-chromosomal STR haplotypes was established for forensic purposes [3–6].

While the Finnish population is known to be genetically homogeneous [7–10] Y STR data on Finns has been lacking. We therefore analysed 400 unrelated Finnish male samples using nine Y-chromosomal STR loci (minimal haplotype), with 200 of these being analysed with an additional seven recently reported [11,12] polymorphic Y-chromosomal STR loci. The gene diversity and exclusion power of the haplotype and individual loci were calculated to assess their usefulness in forensic analysis and paternity testing. Furthermore, we analysed the geographical distribution of

* Corresponding author. Tel.: +358-9-191-27472;

fax: +358-9-191-27518.

E-mail address: antti.sajantila@helsinki.fi (A. Sajantila).

the Y STR haplotypes in a sample set of six subpopulations ($n = 200$), with the origin of the sample being verified two generations back in the male lineage. The subpopulations were compared with the whole data set as well as with each other to evaluate the level of substructure in Finland.

2. Materials and methods

For population genetic analysis the total number of samples from unrelated Finnish males descended from various parts of the country was 400. Two hundred of these samples were obtained from the large DNA registry of the Finnish Red Cross Blood Transfusion Service with the informed consent from individuals of known geographical origin and family history. These 200 samples were classified into six geographical regions of sample origin (Häme $n = 30$, Karjala $n = 30$, Pohjanmaa $n = 30$, Satakunta $n = 30$, Savo $n = 30$ and Varsinais-Suomi $n = 30$; 20 samples had their origin in the central or northern parts of Finland). Blood samples were collected from only those males known to be unrelated in recent family history and who were representatives of only the same generation (aged between 40 and 55 years). DNA was extracted from lymphocytes by standard methods [13].

For paternity evaluation, we analysed 400 samples from the Paternity Testing Laboratory, National Public Health Institute, Finland. Samples were from 200 routine paternity cases (alleged father and son), where 100 were confirmed with nine to fifteen autosomal STR loci to be biological father–son pairs (PI over 99.8%) and 100 were exclusion cases (three or more non-matching loci). The alleged father samples ($n = 200$) were included in further population analyses. DNA was extracted from lymphocytes using the Chelex method.

The loci were amplified in two multiplex PCRs, a 9-plex [4] and a 10-plex [14] reaction, with three overlapping Y STR loci (DYS19, DYS391, DYS392). Nine of these 16 loci belong to the widely used Y-chromosomal minimal haplotype [15]. The seven additional loci were DYS435, DYS436, DYS437, DYS438, DYS439, DYS460 (a.k.a. Y GATA A7.1) and Y GATA H4. The PCR-products were subsequently analysed using capillary electrophoresis (ABI 310, Applied Biosystems). Reference individuals with known repeat unit lengths for the markers were run alongside the population samples in every run. A total of 400 unrelated Finnish males were analysed using minimal haplotype and 200 of them were analysed using 16 Y STR loci. The population data from the minimal haplotype were submitted to the European Y STR haplotype reference database [4]. The 16-loci haplotype data is available from corresponding author (AS) upon request.

Average genetic diversity (GD) and comparison of population samples (F_{st}) were calculated using Arlequin software, version 2.0 [16]. Power of exclusion (PE) was calculated as $1 - \sum P_i^2$, where P_i is the frequency of the haplotype.

3. Results and discussion

3.1. Finnish Y-chromosomal population structure

Using the set of minimal haplotype markers, our population sample ($n = 400$) consisted of 146 different haplotypes, 107 (73.3%) of which occurred only once in our population sample. One haplotype was strikingly frequent, observed in 90 (22.5%) samples (Fig. 1a). This haplotype was observed 14 times in the European reference database [4] in among total twelve 802 samples, all in the populations surrounding Finland. Of the five most common Finnish haplotypes, all of them were also observed elsewhere in Europe [4]. In the US reference database [5], only one (14-12-28-23-10-11-13-14-14) of these five Finnish haplotypes was observed, with frequency of four out of 1705 samples. In the Asian reference database [6], also only one of these haplotypes (14-13-29-23-10-14-14-11-13) was present, occurring with a frequency of one out of 2387 samples. Moreover, these five commonly found haplotypes were observed in other Finno-Ugric speaking populations in Europe with a frequency up to 16.0% (Pimenoff et al., unpublished data).

When the number of observed haplotypes was divided by the number of samples, the fraction reflecting heterogeneity in the Finnish population was small (36.5%) compared with that found on average in other European populations (83.7%) [4]. In the surrounding populations (Estonia, Russia (Novgorod and Moscow), Norway and Sweden), this value varies between 63 and 96% [4], indicating a local population bottleneck in Finland.

Two hundred samples were further analysed with the 10-plex reaction. In the 16-loci haplotype data, constructed from the minimal haplotype and the seven additional loci from the 10-plex reaction, 120 different haplotypes were found. Of these haplotypes, 99 (82.5%) occurred only once (Fig. 1b). The most common haplotype was shared by 13.0% of the samples. Although the fraction of different haplotypes rose to 60.0% using 16 loci, this was still much smaller than fraction obtained on average in other European populations from minimal haplotype data.

The number of alleles observed in each locus varied from 2 (DYS435) to 13 (DYS385) (Table 1). Genetic diversity values of separate microsatellites ranged in minimal haplotype loci between 0.316 (DYS19) and 0.721 (DYS385, when two alleles, a and b, were treated separately, and 0.668 when treated as a unit), and in the seven additional loci between 0.020 (DYS436) and 0.526 (Y GATA H4). GD values are shown in Table 1. When haplotypes were constructed beginning with the most variable loci and ending with the least variable loci, and vice versa, it showed that beyond the 10 most informative loci, little resolution was gained by adding other loci (Fig. 2).

We divided our samples according to the geographical region in which sample donors had family roots. This resulted in six subpopulations (Häme, Karjala, Pohjanmaa, Satakunta, Savo and Varsinais-Suomi, Fig. 3a), each con-

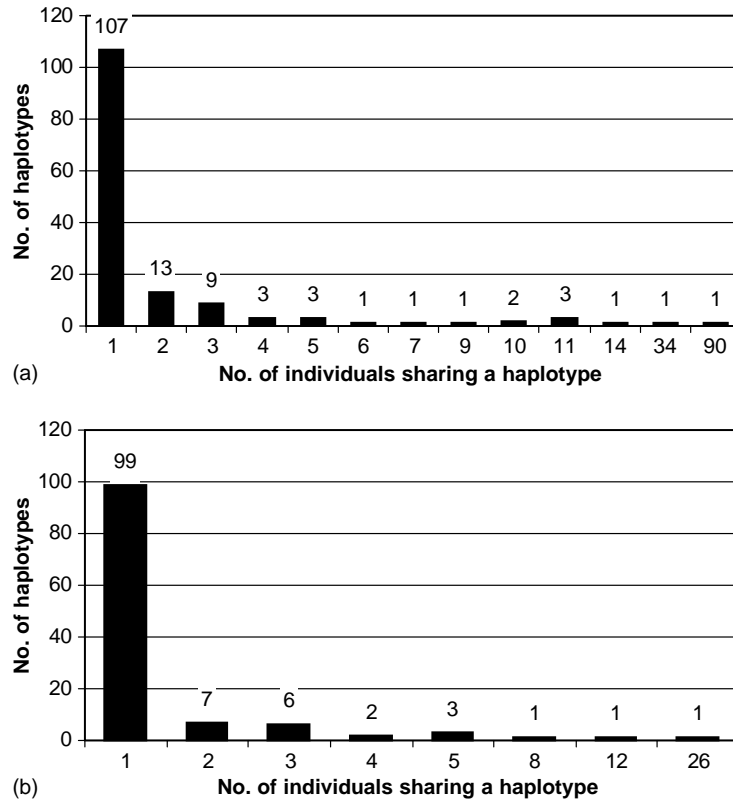


Fig. 1. (a) Minimal haplotype distribution in 400 Finns; (b) 16-loci haplotype distribution in 200 Finns.

taining 30 individuals; a further twenty samples originated from the central and northern parts of Finland. The most common minimal haplotype was found in 26.5% of this sample ($n = 200$). The haplotype occurred with extremely high frequency in Savo (53.3%) and high frequency in Pohjanmaa (43.3%) and Karjala (36.7%) but was absent in Satakunta (Fig. 3a). The most common 16-loci haplotype (13.5%) in Finns was found in four of these subpopulations

(33.3% in Savo, 23.3% in Pohjanmaa, 10.0% in Karjala and 6.7% in Häme); however, in the subpopulation located in south-western Finland (Satakunta and Varsinais-Suomi), this haplotype was not observed at all, and no other haplotype was found to clearly predominate (Fig. 3b). Considering all individual loci, GD values were the lowest among individuals from Savo, with the exception of locus DYS435. Similarly, GD values calculated using minimal and 16-loci

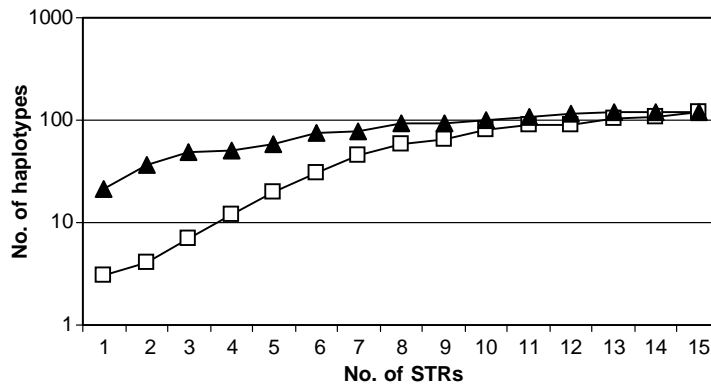


Fig. 2. Number of haplotypes produced by combining most variable loci to least variable loci and vice versa. Data for 200 Finnish males, 16-loci haplotype, DYS385 data combined.

Table 1
Allele frequencies in a Finnish population sample

Allele	DYS19	DYS389	DYS390	DYS391	DYS392	DYS393	DYS385	DYS435	DYS436	DYS437	DYS438	DYS439	DYS460	GATA H4	Alleles	DYS385
7															9, 13	0.003
8											0.005		0.010		10, 13	0.008
9				0.005			0.001				0.015	0.010	0.245	0.005	10, 14	0.003
10				0.428			0.006				0.900	0.705	0.685	0.285	10, 17	0.003
10.2							0.001								10.2, 13	0.003
11		0.005		0.555	0.315		0.338	0.980	0.005		0.055	0.235	0.055	0.625	11, 11	0.003
12		0.218		0.013	0.008	0.015	0.019	0.020	0.990		0.020	0.030	0.005	0.070	11, 12	0.008
13	0.010	0.320			0.038	0.363	0.319		0.005		0.005	0.020		0.015	11, 13	0.545
14	0.818	0.450			0.633	0.588	0.248			0.755					11, 14	0.103
15	0.118	0.008			0.008	0.035	0.054			0.030					11, 15	0.015
16	0.043						0.004			0.205					12, 13	0.018
17	0.013						0.004			0.010					12, 14	0.008
18							0.005								12, 15	0.005
19							0.001								13, 13	0.008
20															13, 14	0.040
21			0.003				0.001								13, 15	0.005
22			0.048												13, 21	0.003
23			0.415												14, 14	0.145
24			0.468												14, 15	0.050
25			0.060												14, 16	0.003
26			0.008												15, 15	0.010
27		0.008													15, 16	0.003
28		0.203													15, 17	0.003
29		0.253													15, 18	0.005
30		0.473													15, 19	0.003
31		0.053													16, 18	0.003
32		0.013													17, 18	0.003
GD	0.316	0.649 0.671	0.605	0.510	0.501	0.523	0.721	0.039	0.020	0.389	0.187	0.449	0.464	0.526		0.668

Data for the first seven loci are derived from 400 samples, for the latter seven are from 200 samples. Locus DYS385 appears twice in the table, first with individual alleles and second with allele combinations.

haplotypes were lowest in the Savo region, indicating a marked reduction in the local male gene pool (Table 2).

Fst analysis revealed a significant substructure in Finland (Table 3). The Savo subpopulation differed significantly from all other subpopulations, and each of the remaining subgroups differed from at least three other subgroups. When comparing the individual subpopulations with the general Finnish database ($n = 200$), we found that the Savo and Satakunta population samples exhibited significant different Fst P -values.

3.2. Paternity testing samples

Power of exclusion (PE), calculated from the minimal haplotype frequencies of 400 unrelated Finnish males, was 92.3%. When seven more loci from the 10-plex reaction were added to construct a 16-loci haplotype, PE reached 97.1% for our population sample of 200 Finnish males.

For paternity testing, we typed the minimal haplotype for 200 alleged father–son pairs; 100 were proved to be true biological father–son pairs based on 9–15 autosomal markers (PI over 99.8%) and 100 were excluded based on three or more autosomal markers. Of the true father–son pairs, no differences between Y-chromosomal haplotypes were found.

Of the excluded father–son pairs, 92 out of 100 had one or more differences in their haplotypes. Alarmingly, eight pairs (8%) could not be excluded due to identical haplotype, and 20 pairs (20%) had two or less excluded Y STRs (Fig. 4). This is in accordance with PE analysis of the Finnish population overall and emphasises the importance of careful population-based assessment of Y STR loci prior to applying analysis of disputed paternity to other male lineages.

Our findings warrant applying a general population database in forensics and paternity testing, particularly in populations, such as in the Finns, where local founder effects have created a substructure. This substructuring is especially apparent when analysing Y-chromosomal markers, which have four times smaller effective population size than autosomal loci and are thus more prone to genetic drift effects [17]. The observed reduction of Y-chromosomal diversity in eastern and northern Finland is not, however, an obstacle for using Y-chromosomal markers in paternity testing. Firstly, movement from rural to southern, urbanised areas started relatively late (in the 1960s) in Finland [7] and this is currently rapidly decreasing the population substructures (Hedman et al., unpublished data). Secondly, our DNA sampling and geographical origin of the population sample were designed to reveal paternal lineage two generations

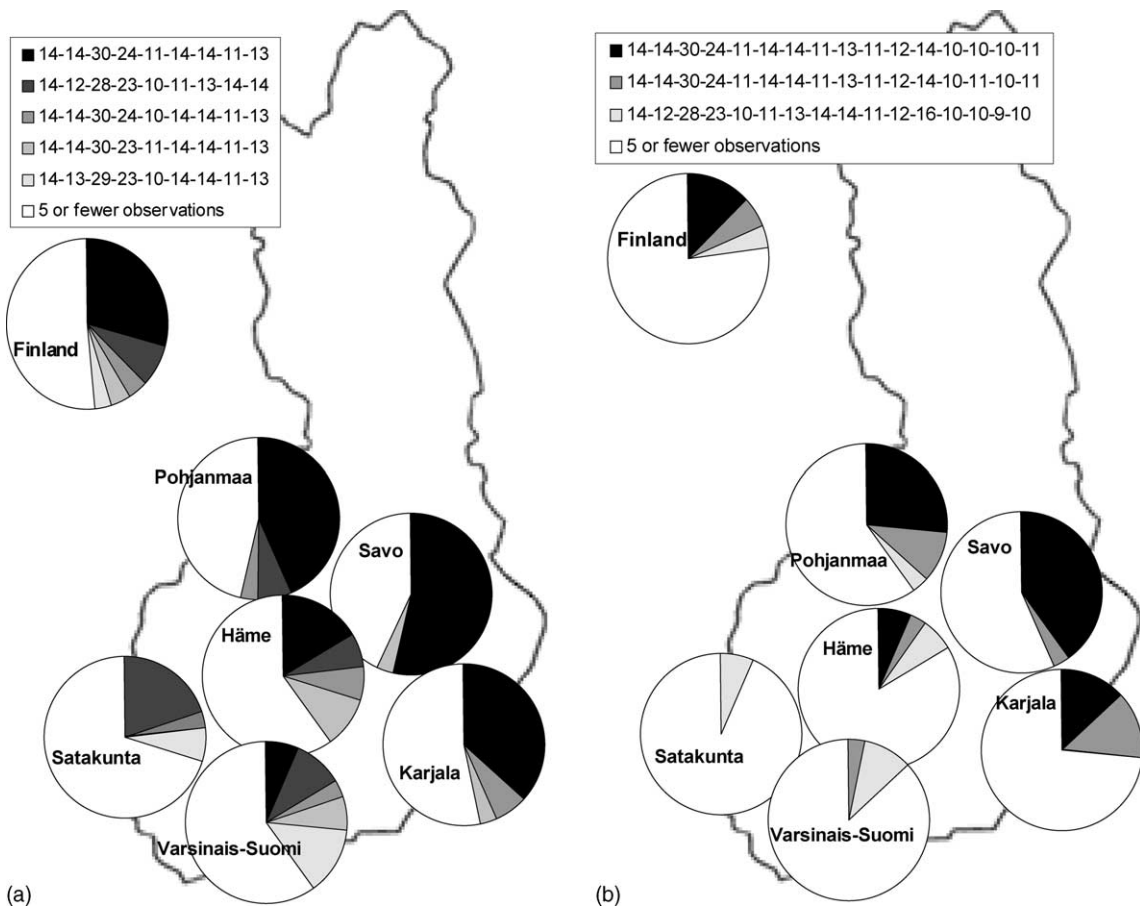


Fig. 3. (a) Distribution of the most common minimal haplotypes in Finnish subpopulations ($n = 200$). (b) Distribution of the most common 16-loci haplotypes in Finnish subpopulations ($n = 200$). Haplotypes constructed from the Y STRs in the following order: DYS19–DYS389I–DYS389II–DYS390–DYS391–DYS392–DYS393–DYS385A–DYS385B–DYS435–DYS436–DYS437–DYS438–DYS439–DYS460–Y GATA H4.

back, which represents Y-chromosomal lineages prior to the current population admixture within Finland. To achieve a higher PE, we suggest using additional loci (e.g. [18–20]) with a minimal haplotype in populations where substructuring is observed.

In conclusion, using a minimal haplotype and seven additional loci, a high resolution Y-chromosomal DNA profile can be achieved for criminal investigations or for establishing biological family history. From a population

history perspective, our data indicate a small male founder population, which is in agreement with earlier studies [9,10,21]. Our findings suggest the occurrence of a population bottleneck, which is demonstrated by reduced Y-chromosomal heterogeneity in the Finns compared to the other European populations on average, and which particularly affected the settlement of the eastern part of Finland in prehistoric times. However, for estimation of the timing and the number of the Y-chromosomal lineages that were

Table 2
Gene diversity values of individual loci, and both minimal and 16-loci haplotypes in different subpopulations

Subpopulation	DYS 19	DYS 389I	DYS 389II	DYS 390	DYS 391	DYS 392	DYS 393	DYS 385	DYS 435	DYS 436	DYS 437	DYS 438	DYS 439	DYS 460	GATA H4	Minimal haplotype	16-loci haplotype
Varsinais-Suomi	0.402	0.669	0.673	0.384	0.536	0.504	0.480	0.751	0.066	0.000	0.370	0.128	0.352	0.577	0.480	0.972	0.986
Pohjanmaa	0.384	0.619	0.549	0.563	0.480	0.434	0.508	0.730	0.000	0.000	0.370	0.248	0.513	0.421	0.545	0.816	0.940
Savo	0.128	0.342	0.434	0.342	0.246	0.131	0.246	0.637	0.066	0.000	0.066	0.000	0.246	0.301	0.487	0.722	0.894
Karjala	0.444	0.536	0.469	0.575	0.453	0.288	0.370	0.651	0.000	0.000	0.301	0.250	0.545	0.306	0.397	0.867	0.956
Satakunta	0.540	0.646	0.683	0.591	0.331	0.581	0.459	0.751	0.128	0.066	0.596	0.402	0.521	0.515	0.545	0.956	0.984
Häme	0.190	0.690	0.720	0.604	0.497	0.521	0.536	0.731	0.000	0.066	0.524	0.193	0.444	0.480	0.600	0.965	0.986

Table 3
Fst *P*-values

	Häme	Karjala	Pohjanmaa	Satakunta	Savo	Varsinais-S	Finns
Häme	–						
Karjala	0.02703*	–					
Pohjanmaa	0.27027	0.47748	–				
Satakunta	0.01802*	0.00000*	0.00901*	–			
Savo	0.00000*	0.01802*	0.01802*	0.00000*	–		
Varsinais-S	0.44144	0.00901*	0.02703*	0.05405	0.00000*	–	
Finns	0.66667	0.08108	0.82883	0.00000*	0.00000*	0.15315	–

Pair-wise comparison between subpopulations and the Finnish population data ($n = 200$). An asterisk (*) denotes *P*-values that are lower than 0.05, indicating a statistical difference between subpopulations. Varsinais-Suomi is abbreviated as Varsinais-S.

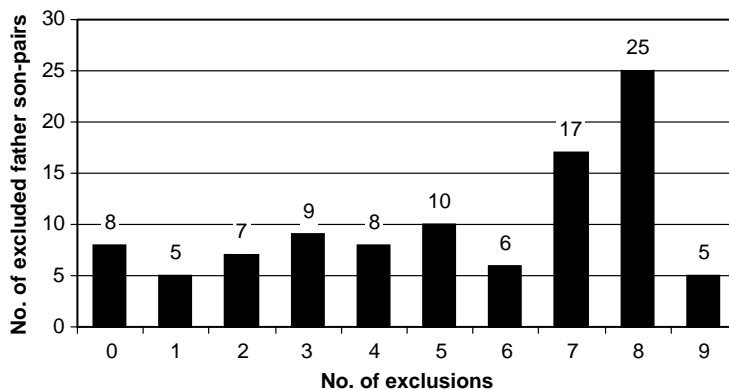


Fig. 4. Experimentally observed number of exclusions in minimal haplotype in 100 excluded father–son pairs. Exclusions were based on analysis of 9–15 autosomal markers with three or more non-matching loci.

involved in this process, further Y-chromosomal analysis including SNPs should be performed.

Acknowledgements

We are grateful to all sample donors. We thank Maikki Latto from the Paternity Laboratory, National Public Health Institute for analysing all of the paternity cases; Kirsti Höök, Eve Karvinen and Teija Vanhala from the Laboratory of Forensic Biology, for help with laboratory experiments; and Professors Antti Penttilä and Erkki Vuori, for continuous support and interest in this study. This study was supported by the European Science Foundation Collaborative Research Programme (OMLL) and Finnish Academy (grant no. 54938).

References

- [1] M.A. Jobling, C. Tyler-Smith, Fathers and sons: the Y chromosome and human evolution, *Trends Genet.* 11 (1995) 449–456.
- [2] M. Kayser, L. Roewer, M. Hedman, L. Henke, J. Henke, S. Brauer, C. Kruger, M. Krawczak, M. Nagy, T. Dobosz, R. Szibor, P. de Knijff, M. Stoneking, A. Sajantila, Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs, *Am. J. Hum. Genet.* 66 (2000) 1580–1588.
- [3] L. Roewer, M. Krawczak, S. Willuweit, M. Nagy, C. Alves, A. Amorim, K. Anslinger, C. Augustin, A. Betz, E. Bosch, A. Cagliá, A. Carracedo, D. Corach, A.-F. Dekairelle, T. Dobosz, B.M. Dupuy, S. Füredi, C. Gehring, L. Gusmaõ, J. Henke, L. Henke, M. Hidding, C. Hohoff, B. Hoste, M.A. Jobling, H.J. Kärger, P. de Knijff, R. Lessing, E. Liebeherr, M. Lorente, B. Martínez-Jarreta, P. Nievas, M. Nowak, W. Parson, V.L. Pascali, G. Penacino, R. Ploski, B. Rolf, A. Sala, U. Schmidt, C. Schmitt, P.M. Schneider, R. Szibor, J. Teifel-Greding, M. Kayser, Online database of European Y-chromosomal short tandem repeat (STR) haplotypes, *Forensic Sci. Int.* 118 (2001) 106–113.
- [4] European Y-STR Haplotype Reference Database (<http://www.ystr.org/>).
- [5] Y-STR Haplotype Reference Database for US Populations (<http://www.ystr.org/usa>).
- [6] Y-STR Haplotype Reference Database for Asian Populations (<http://www.ystr.org/asia>).
- [7] H.R. Nevanlinna, The Finnish population structure. A genetic and genealogical study, *Hereditas* 71 (1972) 195–236.
- [8] R. Norio, H.R. Nevanlinna, J. Perheentupa, Hereditary diseases in Finland; rare flora in rare soil, *Ann. Clin. Res.* 5 (1973) 109–141.

- [9] A. Sajantila, A.-H. Salem, P. Savolainen, K. Bauer, C. Gierig, S. Pääbo, Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population, *Proc. Natl. Acad. Sci. U.S.A.* 93 (1996) 12035–12039.
- [10] P. Lahermo, M.L. Savontaus, P. Sistonen, J. Beres, P. de Knijff, P. Aula, A. Sajantila, Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami, *Eur. J. Hum. Genet.* 7 (1999) 447–458.
- [11] P.S. White, O.L. Tatum, L.L. Deaven, J.L. Longmire, New, male-specific microsatellite markers from the human Y chromosome, *Genomics* 57 (1999) 433–437.
- [12] Q. Ayub, A. Mohyuddin, R. Qamar, K. Mazhar, T. Zerjal, S.Q. Medhi, C. Tyler-Smith, Identification and characterisation of novel human Y-chromosomal microsatellites from sequence based information, *Nucleic Acids Res.* 28 (2000) e8.
- [13] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic Acid Res.* 16 (1988) 1215.
- [14] C.M. Ruitberg, J.M. Butler, New primer sets for Y chromosome and CODIS STR loci, in: *Proceedings of the Poster Presented at International Symposium of Human Identification*, Biloxi, MS, USA, 2000.
- [15] L. Roewer, M. Kayser, P. de Knijff, K. Anslinger, A. Betz, A. Caglià, D. Corach, S. Füredi, L. Henke, M. Hidding, H.J. Kärigel, R. Lessig, M. Nagy, V.L. Pascali, W. Parson, B. Rolf, C. Schmitt, R. Szibor, J. Teifel-Greding, M. Krawczak, A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males, *Forensic Sci. Int.* 114 (2000) 31–43.
- [16] S. Schneider, D. Roessli, L. Excoffier, Arlequin ver 2.000, 2000 (<http://www.anthro.unige.ch/arlequin>).
- [17] P. de Knijff, Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome, *Am. J. Hum. Genet.* 67 (2000) 1055–1061.
- [18] J.M. Butler, R. Schoske, P.M. Vallone, M.C. Kline, A.J. Redd, M.F. Hammer, A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers, *Forensic Sci. Int.* 129 (2002) 10–24.
- [19] A.J. Redd, A.B. Agellon, V.A. Kearney, V.A. Contreras, T. Karafet, H. Park, P. de Knijff, J.M. Butler, M.F. Hammer, Forensic value of 14 novel STRs on human Y chromosome, *Forensic Sci. Int.* 130 (2002) 97–111.
- [20] E. Bosch, A.C. Lee, F. Calafell, E. Arroyo, P. Henneman, P. de Knijff, M.A. Jobling, High resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions, *Forensic Sci. Int.* 125 (2002) 42–51.
- [21] R.A. Kittles, M. Perola, L. Peltonen, A.W. Bergen, R.A. Aragon, M. Virkkunen, M. Linnoila, D. Goldman, J.C. Long, Dual origins of Finns revealed by Y chromosome haplotype variation, *Am. J. Hum. Genet.* 62 (1998) 1171–1179.