



Y-Chromosomal SNPs in Finno–Ugric-Speaking Populations Analyzed by Minisequencing on Microarrays

Mirja Raitio, Katarina Lindroos, Minna Laukkanen, et al.

Genome Res. 2001 11: 471-482

Access the most recent version at doi:[10.1101/gr.156301](https://doi.org/10.1101/gr.156301)

References This article cites 41 articles, 15 of which can be accessed free at:
<http://genome.cshlp.org/content/11/3/471.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center is a white box with the text "LEARN MORE". On the right is a woman wearing a red and white superhero cape and mask, with the Cellecta logo (a green molecular structure) and the word "CELLECTA" below it.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Methods

Y-Chromosomal SNPs in Finno-Ugric-Speaking Populations Analyzed by Minisequencing on Microarrays

Mirja Raitio,^{1,5} Katarina Lindroos,^{1,4,5} Minna Laukkanen,¹ Tomi Pastinen,¹ Pertti Sistonen,² Antti Sajantila,³ and Ann-Christine Syvänen^{1,4,6}

¹Department of Human Molecular Genetics, National Public Health Institute, Helsinki, Finland; ²Finnish Red Cross Blood Transfusion Service, Helsinki, Finland; ³Department of Forensic Medicine, University of Helsinki, Helsinki, Finland; ⁴Molecular Medicine, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

An increasing number of single nucleotide polymorphisms (SNPs) on the Y chromosome are being identified. To utilize the full potential of the SNP markers in population genetic studies, new genotyping methods with high throughput are required. We describe a microarray system based on the minisequencing single nucleotide primer extension principle for multiplex genotyping of Y-chromosomal SNP markers. The system was applied for screening a panel of 25 Y-chromosomal SNPs in a unique collection of samples representing five Finno-Ugric populations. The specific minisequencing reaction provides 5-fold to infinite discrimination between the Y-chromosomal genotypes, and the microarray format of the system allows parallel and simultaneous analysis of large numbers of SNPs and samples. In addition to the SNP markers, five Y-chromosomal microsatellite loci were typed. Altogether 10,000 genotypes were generated to assess the genetic diversity in these population samples. Six of the 25 SNP markers (*M9*, *Tat*, *SRY1083I*, *M17*, *M12*, *92R7*) were polymorphic in the analyzed populations, yielding six distinct SNP haplotypes. The microsatellite data were used to study the genetic structure of two major SNP haplotypes in the Finns and the Saami in more detail. We found that the most common haplotypes are shared between the Finns and the Saami, and that the SNP haplotypes show regional differences within the Finns and the Saami, which supports the hypothesis of two separate settlement waves to Finland.

The sequence variation on the Y chromosome has appeared to be low, and until recently only a modest number of single nucleotide polymorphisms (SNPs) were known (for references, see Table 1). By screening 18 kb of coding sequence in four genes in the nonrecombining parts of the Y chromosome in samples from the five continents, ~100 new Y-chromosomal SNPs were recently discovered (Shen et al. 2000). The frequency of SNPs was found to be lower than that of autosomal SNPs (Shen et al. 2000), in accordance with the four-times-lower effective population size of the Y chromosome and a recent common ancestor for the Y chromosome (Hammer 1995; Thomson et al. 2000). The Y-chromosomal SNPs are useful biallelic markers for following paternal lineages as a complement to the widely analyzed sequence variation of the maternally inherited mitochondrial DNA (Di Rienzo and Wilson 1991; Sajantila et al. 1995; Sigurðardóttir et al. 2000). Because SNPs in the nonrecombining parts of the Y chromosome can be considered as the results of unique

events during evolution, they can be used in combination with the more rapidly evolving microsatellite markers to construct well-defined haplotypes to improve the power of the statistical analysis of the results (Jobling et al. 1997).

A variety of techniques have been applied for discovering and genotyping Y-chromosomal SNPs; of these, heteroduplex analysis using denaturing high-performance liquid chromatography (DHPLC) has been the most successful approach (Underhill et al. 1997). DHPLC has proven to be particularly efficient for discovering new SNPs, but the technique is useful also for scoring previously known ones (Underhill et al. 1997; Shen et al. 2000). As the number of known SNPs on the Y chromosome increases, methods with higher throughput than those currently available will be required to utilize the potential of known SNP markers in population studies effectively. Microarray-based genotyping technology is a promising alternative for parallel and simultaneous analysis of many SNPs in large sample sets (Hacia 1999; Pastinen et al. 2000). In the present study we describe a microarray system based on the minisequencing single nucleotide primer extension principle (Syvänen et al. 1990; Pastinen et al. 1997) to facilitate rapid and reliable multiplex genotyping of Y-chromosomal SNPs. Using this newly de-

⁵These authors contributed equally to this work.

⁶Corresponding author.

E-MAIL Ann-Christine.Syvanen@medsci.uu.se; FAX 46-18-6112519.
Article and publication are at www.genome.org/cgi/doi/10.1101/gr.156301.

Table 1. PCR and Minisequencing Primers for the Y-Chromosomal SNP Markers

Locus	Primer ^{b,c}	Sequence, 5' to 3' ^d	Size of PCR product (bp)	Nucleotide variation ^e	References
DYS271 ^a (M2)	5' PCR primer 3' PCR primer ms primer	GCT CCC CTG TTT AAA AAT GTA GG TAC AGC TCC CCC TTT ATC CTC C CCT TTA TCC TCC ACA GAT CTC A	136	A/G	Seielstad et al. 1994
92R7 ^a	5' PCR primer 3' PCR primer ms primer	TGC ATG AAC ACA AAA GAC GTA GCA TTG TTA AAT ATG ACC AGC TTA AAT ATG ACC AGC AAA GAC AA	100	C/T	Mathias et al. 1994 Hurles et al. 1999
SRY4064 ^a	5' PCR primer 3' PCR primer ms primer	CCA GCC CTT CGA GAG GTC AA CAG GAT GGT CTC AAT CTC TTC CTC TTC ACC CTG TGA TCC GCT	106	G/A	Whitfield et al. 1995
SRY9138	5' PCR primer 3' PCR primer ms primer	GTT GAT ATG ATA TTA TAG AGG C ACC TAC ATG CAA AAA TAA GGA TG GAG GCA AAT TTA ATG CTC TCG G	118	C/T	Whitfield et al. 1995
SRY10831 ^a (SRY-1532)	5' PCR primer 3' PCR primer ms primer	TGG CCT CTT GTA TCT GAC TTT TT CAA GGC ACC ACA TAG GTG AAC CAT AGG TGA ACC TTG AAA ATG TTA	112	A/G	Whitfield et al. 1995 Kwok et al. 1996 Veitia et al. 1997
DYS199 (M3)	5' PCR primer 3' PCR primer ms primer	AAT AGG TTG ACC TGA CAA TGG TCA TTT TAG GTA CCA GCT CTT AAT GGG TCA CCT CTG GGA CTG A	113	C/T	Underhill et al. 1996
PN1	5' PCR primer 3' PCR primer ms primer	AGG AAC AAG GGT CTT GAG AGG TAT CTC TTC CCC TTT AAG ACA A AAG GGT CTT GAG AGG GAG AGC	98	C/T	Hammer et al. 1997
PN2 ^a	5' PCR primer 3' PCR primer ms primer	CTT CAT GCT GGT TAA AGG AGC A ACC TCA GAA TCA GCT CCA AGG TGC CCC TAG GAG GAG AA	145	C/T	Hammer et al. 1997
PN3	5' PCR primer 3' PCR primer ms primer	AGC ATC TGA GAA AAC TGC CTC CAA TGC TCC TAA CTC TAA TTT AC TCT GAG AAA ACT GCC TCT GCT TGA	104	A/G	Hammer et al. 1997
Tat ^a	5' PCR primer 3' PCR primer ms primer	CTC TGA GTG TAG ACT TGT GA GAA GGT GCC GTA AAA GTG TGA A CTC TGA AAT ATT AAA TTA AAA CAA C	146	T/C	Zerjal et al. 1997
SRY2627	5' PCR primer 3' PCR primer ms primer	CGC AGT TTG GCT TCC GGG CC TTC ACC CTG TAG GGC ATG GG CCA AGG AAG CCC CAC AGG GTG C	121	C/T	Veitia et al. 1997 Bianchi et al. 1997
M4 ^a (DYS234)	5' PCR primer 3' PCR primer ms primer	TCC TAG GTT ATG ATT ACA GAG CG GCA GAA CAT TTG TAC TGT TCC ACT AGA TAT AGA TCT TAG ATA TGA TTA	319	A/G	Underhill et al. 1997
M6	5' PCR primer 3' PCR primer ms primer	ACC ACA TTT CTG GTT GGC TTG T CCC TAC TTT GAT GGC TAA TAC TA TGG TTG GCT TGT AGT TCT TTC T	132	C/T	Underhill et al. 1997
M7	5' PCR primer 3' PCR primer ms primer	GCA TCA CCA AAG GGC ATG TAA TC TGA GTT ACT GTT CTT CTT GTT GTT AAA GGG CAT GTA ATC ATT CCT CT	116	C/G	Underhill et al. 1997
M8 ^a	5' PCR primer 3' PCR primer ms primer	CCC ACC CAC TTC AGT ATG AA ACT CGA GGC TGA CAG AC GAC CAA TAG AAG AAC TGC CCA A	318	G/T	Underhill et al. 1997
M9	5' PCR primer 3' PCR primer ms primer	GTA AGA CAT TGA ACG TTT GAA CA GCA GCA TAT AAA ACT TTC AGG A GAA ACG GCC TAA GAT GGT TGA AT	154	C/G	Underhill et al. 1997
M11	5' PCR primer 3' PCR primer ms primer	GCA TAA ACA AGA ACT TAC TGA G TAT CTC TCT GTC TGT CTC TC CCC TCT CTC CTT GTA TTC TAA C	270	A/G	Underhill et al. 1997
M12 ^a DYS260	5' PCR primer 3' PCR primer ms primer	CTC CCG AGT AGC TGG GAA AGC AGT CCA AGA CTA GCC TGA G TGA GCA ACA TAG TGA CCC CCA T	155	G/T	Underhill et al. 1997
M13	5' PCR primer 3' PCR primer ms primer	GCC ATG ATT TTA TCC AAC C GGG CAG TAG GTA GGT TAA GGG GTA GGT TAA GGG CAA GAC GGT TA	142	C/G	Underhill et al. 1997
M14	5' PCR primer 3' PCR primer ms primer	TGT TAA AAC TGA GCT TCA TTA AC GGT TGG ATA AAA TCA TGG CT ATT ATA AAT GAA AAA ATC AGT AAC ATG	204	T/C	Underhill et al. 1997
M17	5' PCR primer 3' PCR primer ms primer	CAA ACT GAT GTA GAG ACA TCT ACT CAC CAG AGT TTG TGG TTG GTG GTT GCT GGT TGT TAC GGG	175	G/T ^f	Underhill et al. 1997
M19	5' PCR primer 3' PCR primer	CAA ACT GAT GTA GAG ACA TCT ACT CAC CAG AGT TTG TGG TTG	175	A/T	Underhill et al. 1997

Table 1. (Continued)

Locus	Primer ^{b,c}	Sequence, 5' to 3' ^d	Size of PCR product (bp)	Nucleotide variation ^e	References
M20	ms primer	AAA CTA TTT TTG TGA AGA CTG TTG TA	211	A/G	Underhill et al. 1997
	5' PCR primer	TCT ACA TGT TCA GTG CAA ATG CA			
	3' PCR primer	GAT TGG GTG TCT TCA GTG CT			
M21 ^a	ms primer	GTC AAC CAA CTG TGG ATT GAA AAT	197	A/T	Underhill et al. 1997
	5' PCR primer	ATT TGA GCA GGC AGA AGA ATC A			
	3' PCR primer	TTG TGG GTA ATA GGC ATG TCT CTG GAA GCT AAG TCA ACT GAA AC			
M22 ^a	ms primer	TCC CTA TTT GTT TTA AGG TGC CA	121	A/G	Underhill et al. 1997
	5' PCR primer	GCA GAC TAA GGC CCA TAT GGG			
	3' PCR primer	TAT GGG CAC TGG GGG AAG GAG G			

^aDetection from complementary DNA strand compared to reference.

^bPCR primers contained the nonspecific universal sequences TTC TAA TAC GAC TCA CTA TAG GGA GA and GCG GTC CCA AAA GGG TCA GT in the 5' end of either the forward or the reverse primers. The minisequencing primers contained a spacer-sequence of 15 T-residues in their 5' ends.

^cms = Minisequencing.

^dPrimer sequences shown in bold were designed for this study; unbolded primer sequences are from the reference.

^eThe nucleotide variation is given as in the reference.

^fM17 is a deletion of G; T is the next nucleotide detected in the minisequencing reaction.

veloped genotyping system we screened a panel of 25 human Y-chromosomal SNPs in a unique collection of 300 samples representing five Finno-Ugric-speaking populations. The samples were from Finns originating from various geographical locations in Finland, three different Saami groups, Karelians from Russia, and Ob-Ugric Mansi and Khanty speakers (Fig. 1). In addition to the Y-chromosomal SNPs, we typed five Y-chromosomal microsatellite loci in these samples. Al-

together, 10,000 genotypes were generated to assess the genetic diversity in these population samples.

RESULTS AND DISCUSSION

The Genotyping System

A method based on minisequencing single nucleotide primer extension (Syvänen 1999) in a micorarray format was developed for multiplex genotyping of 25 Y-

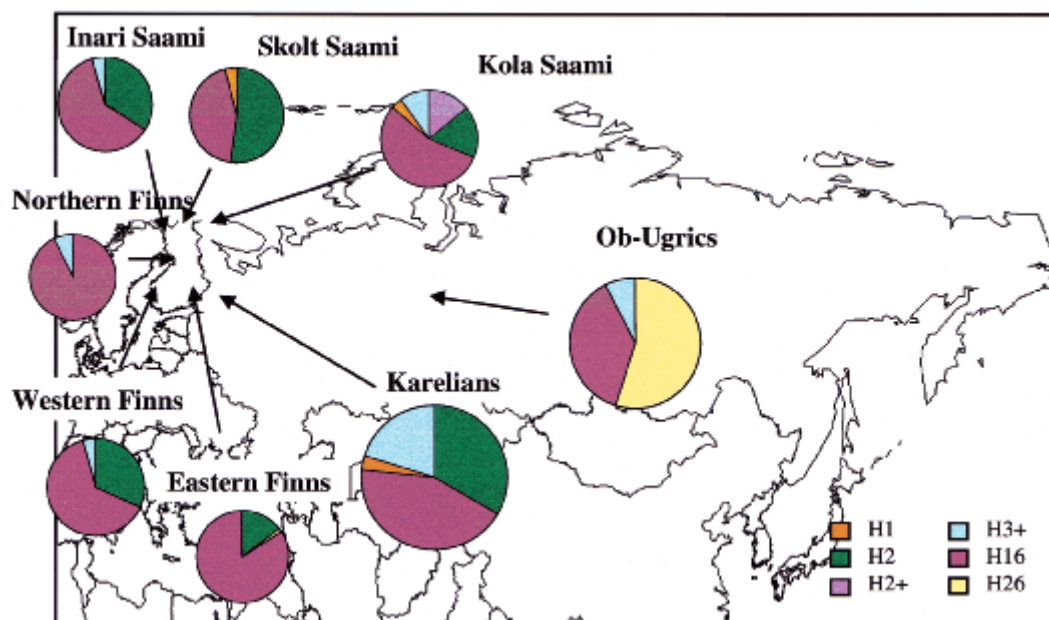


Figure 1 The geographic origin of the sampled Finno-Ugric-speaking populations and the frequencies of Y-chromosomal haplotypes formed by SNPs in these populations. For haplotype designation and nomenclature as well as the population frequencies of the observed haplotypes, see Table 5.

chromosomal SNPs. The 25 SNPs to be included in the panel were selected from the literature, and mini-sequencing primers for each SNP were designed to anneal to the DNA region immediately adjacent to the site of the SNP (Table 1). The primers were immobilized covalently on glass microscope slides as 5×5 arrays (Fig. 2). The genotyping procedure involves (1) multiplex PCR amplification of the Y-chromosomal DNA regions spanning the SNPs; (2) extension of the immobilized primers with labeled ddNTPs using a DNA polymerase with the multiplex PCR products as templates; (3) measurement of the incorporated label on the microarrays; and (4) interpretation of the results.

Performing multiplex PCR amplification is a major rate-limiting step in all currently applied genotyping procedures. Amplification of more than 10 fragments reproducibly and successfully from multiple samples has proven to be difficult (Hacia et al. 1998; Pastinen et al. 2000). In the present study a touchdown PCR procedure (Don et al. 1991) was used to circumvent differences in amplification efficiency arising from differences in melting temperatures between the PCR primers in combination with universal 5' sequences on the primers (Shuber et al. 1995). This procedure allowed us to unify the reaction kinetics of the primer annealing and amplify 24 Y-chromosomal SNPs as two sets of multiplex PCRs with 17 and 6 primer pairs, respectively. Although the success rate of the optimized multiplex PCRs was 100% for six of the SNPs, in 4% of the samples one or more SNP sites failed to amplify in the multiplex PCR (Table 2). These failures were recognized as absence of signal in the reactions on the microar-

Table 2. Success Rate of Multiplex PCRs

SNP-marker	Success rate (%)
Tat	100
PN3	100
M19	100
SRY4064	100
M4	100
SRY10831	98
M17	98
M20	98
SRY2627	96
M9	94
M6	94
M14	90
DYS199	88
M11	85
SRY9138	85
M8	83
M7	60
DYS271	46
6 plex PCR	
PN2	100
M12	96
M21	96
Pn1	88
M22	87
M13	81

^aThe M17 and M19 SNP sites are amplified within one PCR fragment.

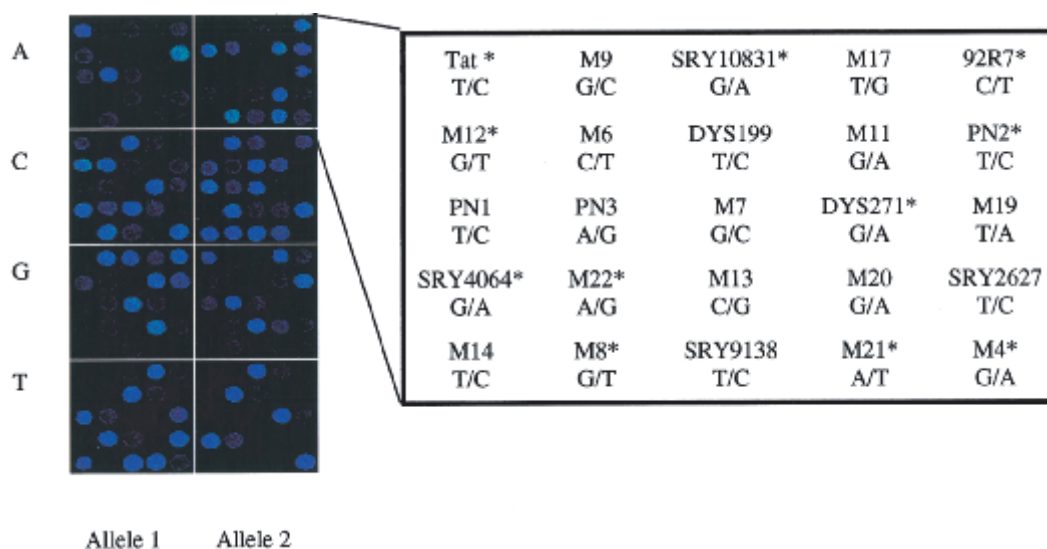


Figure 2 Result from validation of the microarray-based genotyping system. Two mixtures of 25 synthetic oligonucleotide templates corresponding to one of the alleles at each SNP site were analyzed in minisequencing reactions on the microarrays to verify that both alleles of each SNP marker would be detected. (*Left panel*) The fluorescence image obtained from four detection reactions (A, C, G, T) with TAMRA-labeled ddNTPs. (*Right panel*) Order of the SNP markers on the array with the nucleotide variation at each site. The asterisk indicates those alleles that are detected from complementary DNA strands compared to earlier published literature.

rays, and these genotypes were obtained after reamplification of the first PCR product, followed by retyping using microarrays. The marker 92R7 was amplified and genotyped individually by minisequencing in the microtiter plate format (Syvänen 1997) because the SNP is located in one out of several repetitive segments (C. Tyler-Smith, pers. comm.), and our initial experiments showed that it was difficult to genotype on the arrays. The minisequencing reactions were optimized with respect to concentration of detection primers during their immobilization to the microarrays, and with respect to concentration of ddNTPs and DNA polymerase during the reactions on the microarrays. The high reaction temperature and short reaction time were found to be critical for avoiding false positive signals caused by template-independent extension of some of the primers. The protocol given in the Methods section is the result of these optimization experiments.

To validate that the microarray-based system would detect either allele of all 25 Y-chromosomal SNPs correctly, two mixtures of artificial single-stranded oligonucleotide templates representing one of the alleles at each site were analyzed. Figure 2 shows

the fluorescence image from this analysis. Table 3 gives the corresponding numeric values measured by the fluorescence scanner and the calculated signal ratios defining the haploid Y-chromosomal genotypes. As can be seen, the 5-fold to infinite differences in signal ratios (R values) between the two genotypes provide unequivocal results at all sites. Unequivocal discrimination between genotypes is achieved also when multiplex PCR fragments amplified from the genomic DNA samples are analyzed. The range of R values for the monomorphic markers was 0.005–0.070 with an average of 0.03, or 0.98–1.0 with an average of 0.99, respectively. Table 4 illustrates the genotyping results when 5 or 10 samples of each genotype at the polymorphic *Tat*, *M9*, *SRY10831*, and *M17* SNP sites were analyzed using both radioactive (^{33}P) and fluorescence (TAMRA) detection. Both the detection systems perform well, with the largest difference in fluorescent and radioactive signal ratios for the marker *M9*. The microarray-based minisequencing system allows genotyping of diploid SNPs both using ^{33}P as label as shown in a previous study (Pastinen et al. 1998) and using fluorescence based on our more recent work (K. Lindroos et

Table 3. Results from Fluorescent Detection of Both Alleles of 25 Y-Chromosomal SNPs in Artificial Single Stranded Templates

SNP (Allele1/Allele2)	Mixture of alleles 1			Mixture of alleles 2			Power of genotype discrimination ^d
	signal ^b		R-value ^c	signal ^b		R-value ^c	
	allele 1	allele 2		allele 1	allele 2		
Tat ^a (T/C)	1808	4	1.0	0	1415	0	∞
M9 (G/C)	4673	0	1.0	0	3643	0	∞
SRY10831 ^a (G/A)	3436	5	1.0	496	3481	0.13	7.7
M17 (T/G)	4592	339	0.93	32	4094	0.0078	120
92R7 ^a (C/T)	4882	175	0.97	162	6584	0.024	40
M12 ^a (G/T)	7305	162	0.98	654	5937	0.099	9.9
M6 (C/T)	3871	0	1.0	45	2964	0.015	67
DYS199 (T/C)	3497	11	1.0	22	3950	0.0055	180
M11 (G/A)	4689	0	1.0	45	7835	0.0057	170
PN2 ^a (T/C)	9905	1028	0.91	1440	6348	0.19	4.8
PN1 (T/C)	3779	4	1.0	0	5473	0	∞
PN3 (A/G)	3236	34	1.0	0	2890	0	∞
M7 (G/C)	6559	0	1.0	112	3357	0.032	31
DYS271 ^a (G/A)	3528	0	1.0	0	3454	0	∞
M19 (T/A)	1011	0	1.0	31	1879	0.016	63
SRY4064 ^a (G/A)	4803	2	1.0	0	3622	0	∞
M22 ^a (A/G)	3872	669	0.85	382	5113	0.070	12
M13 (C/G)	3470	0	1.0	96	4397	0.021	48
M20 (G/A)	7076	5	1.0	422	4619	0.084	12
SRY2627 (T/C)	3939	4	1.0	0	4779	0	∞
M14 (T/C)	1815	0	1.0	0	2354	0	∞
M8 ^a (G/T)	6553	83	0.99	2104	9072	0.19	5.2
SRY9138 (T/C)	3131	141	0.96	0	3817	0	∞
M21 ^a (A/T)	4080	4	1.0	0	6681	0	∞
M4 ^a (G/A)	5187	2	1.0	0	3018	0	∞

^aDetection from complementary strand.

^bArbitrary fluorescence units.

^cR-value = Allele1/(Allele 1 + Allele 2).

^d $R_{\text{mix. of alleles 1}}/R_{\text{mix. of alleles 2}}$.

Table 4. Accuracy of Genotype Discrimination and Comparison of Radioactive and Fluorescent Detection in Minisequencing on Microarrays for Four Polymorphic Y-Chromosomal SNPs

Sample genotype ^b	Tat ^a (T/C) R = T/(T+C)		M9 (G/C) R = G/(G+C)		SRY10831 ^a (G/A) R = G/(G+A)		M17 (T/G) R = T/(T+G)	
	T	C	G	C	G	A	T	G
Fluorescent detection								
Variation ^c	0.90–0.93	0–0.07	0.91–1.0	0.06–0.40	0.75–1.0	0–0.04	0.70–0.80	0–0.08
Average	0.93	0.04	0.97	0.16	0.87	0.02	0.74	0.04
Power of genotype discrimination ^d	23		6.1		44		19	
Radioactive detection								
Variation	0.99–1.00	0.00–0.05	0.93–0.99	0–0.09	0.94–1.0	0–0.04	0.92–0.99	0–0.05
Average	1.00	0.02	0.96	0.03	0.96	0.02	0.97	0.02
Power of genotype discrimination ^d	50		32		48		49	

^aDetection from complementary strand.^bData from ten samples of each genotype was used, except for the fluorometric measurements of M9 C allele, SRY10831 G allele and M17 T allele where only five samples were used.^cThe range of variation between the lowest and highest R values.^d $R_{\text{average of alleles1}}/R_{\text{average of alleles2}}$.

al., unpubl.). The major advantage of the fluorescence detection system is the higher resolution of the fluorescence scanner, which allows measurement of an extremely large number of data points on a miniaturized area.

Y-Chromosomal SNP and Haplotype Frequencies

Our analysis of the Finnish, Saami, Karelian, Mansi, and Khantyi samples for the 25 Y-chromosomal SNP markers included in the panel revealed that five of them (*M9*, *Tat*, *SRY10831*, *M17*, and *92R7*) were polymorphic in all our population samples. The SNP *M12* (*DYS260*), with a G → T transversion, was also polymorphic to some extent in the Kola-Saami samples (Fig. 3). Interestingly, the *M12* T allele has also been found in other European populations as well as in sub-continent Indians (Underhill et al. 1997). The rest of the 19 markers were monomorphic, which agrees well with the earlier published literature (Bianchi et al. 1997; Hammer et al. 1997; Karafet et al. 1999). The genotypes of the polymorphic markers were confirmed

by solid-phase minisequencing in a microtiter format (data not shown; Syvänen 1997).

Thus, we found six polymorphic Y-chromosomal SNPs in our population samples, and six different haplotypes could be constructed from these (see Table 5 for the haplotype nomenclature). The most striking difference in haplotype frequencies between the populations is that the haplotype *H26* is present at high frequency in the Ob-Ugric sample (55%, Fig. 1), whereas it is absent in other groups, with the exception of the Finns, where it was found once in the subgroup of Eastern Finns (Fig. 1; Table 5). On the other hand, haplotype *H2*, which occurs with high frequency in other population samples (15%–52%), is absent in our Ob-Ugric sample. Haplotype *H16*, which has earlier been reported to be present in several Northern Eurasian samples, was represented in our study with high frequency in all samples at a frequency varying from 38% in Ob-Ugrics to 77% in the Finns. Figure 1 and Table 5 summarize the haplotype frequencies in the analyzed populations.

Many of the SNPs analyzed in our panel seem to be restricted to specific populations. The *DYS199* C → T transition, for instance, has so far been shown to be polymorphic only among AmerIndians and in the Chukchan populations in Northeastern Siberia (Underhill et al. 1996; Karafet et al. 1997; Lell et al. 1997; Ruiz-Linarez et al. 1999). Only the C allele was identified in our sample collection. Similarly, the SNP *M19*, with a T → A transversion, has been found only in some South American populations (Underhill et al. 1997; Ruiz-Linarez et al. 1999). The *Tat* T → C transition is also geographically restricted and has so far been observed only in Northern Eurasian populations (Zerjal et al. 1997; Lahermo et al. 1999). In our samples the Finns had the highest frequency (78%) of the C

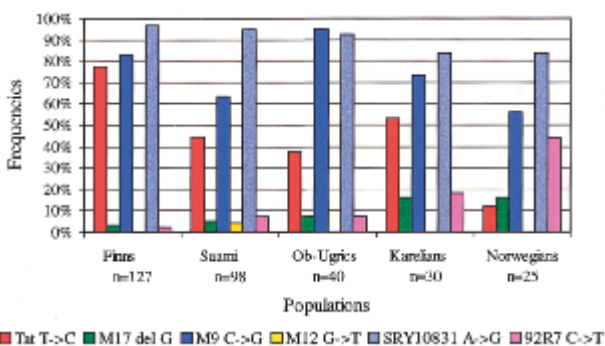
**Figure 3** Allele frequencies of six polymorphic Y-chromosomal SNP markers.

Table 5. Nomenclature^a and Frequencies for Y-Chromosomal SNP Haplotypes^b

Haplotype	Locus	Haplotype frequency (%)												
		All Finns (n = 127)	Northern Finns (n = 14)	Western Finns (n = 47)	Eastern Finns (n = 66)	All Finns (n = 127)	Inari Saami (n = 46)	Kola Saami (n = 29)	Skolt Saami (n = 23)	All Saami (n = 98)	Karelians (n = 30)	Mansi (n = 19)	Khanty (n = 21)	All Ob-Ugric (n = 40)
H2+	T	0	0	0	0	0	0	14	0	4.1	0	0	0	0
H2	T	15	0	32	15	19	34	17	52	33	0	0	0	0
H26	T	1.6	0	0	1.6	0.8	0	0	0	0	0	33	0	0
H16	C	84	93	64	84	77	61	55	44	55	43	57	55	38
H1	C	0	0	0	0	0	0	3.4	4.3	2.0	3.0	0	0	0
H3+	T	0	7.1	4.3	0	2.4	4.5	10	0	5.2	20	5.3	9.5	7.5
	Nucleotide diversity	0.087	0.187	0.194	0.087	0.202	0.202	0.305	0.187	0.339	0.339	0.117	0.192	
	Mean pairwise difference	0.53	0.57	1.16	0.53	1.21	1.21	1.83	1.12	2.04	2.04	0.70	1.15	

^aThe nomenclature used in this study is according to published literature (Jobling and Tyler-Smith 1997; Tyler-Smith 1999).^bThe substitutions from ancestral state are shown bolded. The A-allele at SRY 10831 site is revertant on an M9-G background (Hammer et al. 1998).^cDetection from complementary strand.

allele, but this variant also appeared in all the other populations, including the Norwegians (which were used as a reference population) with a frequency of 12% (Fig. 3). In an earlier study the frequency in Norwegians was found to be 4% (Zerjal et al. 1997). However, the C → G substitution at the *M9* locus appears to be an old mutation because it is frequent worldwide with the exception of Africa. Its absence in Africa might suggest that it occurred initially outside Africa during an early divergence out of Africa. The G allele has been shown to be prevalent in Eurasia, which is consistent with our data (Underhill et al. 1997). It is highly frequent in the Finns and in the Saami, and usually occurs on the same chromosome as the *Tat* C allele. Assuming that the order of mutation is correct, and that no back mutations have occurred at these sites, the two major haplogroups are divided by two mutation steps.

One of the major haplotypes (*H2*) carries a chromosome pool with the *M9* C allele together with the *Tat* T allele, whereas the other major haplotype (*H16*) is the result of two subsequent mutations, wherein first the *M9* C has been substituted with G after which the *Tat* T has been substituted with C. The latter has been shown to be the most frequent allele in Finland (Zerjal et al. 1997; Lahermo et al. 1999). This haplotype is highly frequent, especially in the Northern and Eastern parts of Finland, where its frequency reaches 93% and 84%, respectively (Fig. 1; Table 5). These regional differences between the Eastern and Northern parts compared to Western Finland support the earlier archeological and genetic evidence for two separate settlement waves to Finland (Kittles et al. 1998; Lahermo et al. 1999). Even though *YAP+* is supposed to have an equal time depth as the *M9*, and is present in different European populations as well as Asian populations it, has not been found in the Finns and Saami (Lahermo et al. 1999; Altheide and Hammer 1997). This is consistent with our data, as the *YAP+* and *DYS271* G allele are associated, and only the A allele of *DYS271* was found in our samples. In earlier studies the *PN2* C → T transition was reported to occur after the *SRY4064* G → A transition (Altheide and Hammer 1997) and also after the *DYS271* A → G transition in a *YAP+* haplotype lineage (Hammer et al. 1997). Our data also support this finding, because our samples only contained the ancestral allele of these three markers.

To further characterize the genetic hierarchy of the Finno-Ugric-speaking populations we applied an analysis of molecular variance (*AMOVA*) by grouping the populations according to linguistics into Finnish speakers, Saami speakers, and Ob-Ugric speakers. This division also corresponds to their geographical localities, and thus no further division according to geography could be made. According to the *AMOVA* analysis, the vast majority (83%) of the sequence variation on

the Y chromosome based on the analyzed SNPs occurs within different populations, whereas 10% occurs among the linguistic subgroups, and only 6.5% occurs between populations among linguistic subgroups. Also earlier studies based on autosomal markers have indicated that 80%–90% of the genetic variation in humans occurs within populations (Barbujani et al. 1997). Our study confirms this finding, although a large variation owing to geographical distances is not expected, since our study populations are from areas relatively close to each other. Obviously, more populations with large number of individuals should be studied to further characterize the underlying demographic or cultural factors that have played a role in forming the genetic structure of these rather isolated populations.

Y-Chromosomal Microvariation in the Finnish and the Saami Subgroups

We further studied the genetic structure of the Finns and the Saami, who have proved to be genetically distinct based on mitochondrial DNA sequences (Sajantila et al. 1995) and autosomal markers (Sajantila and Pääbo 1995), but similar based on Y-chromosomal microsatellites (Lahermo et al. 1999). For this analysis, we genotyped five Y-chromosomal microsatellite markers, *DYS389-I*, *DYS389-II*, *DYS391*, *DYS393*, and *DYS19*, and assigned the microsatellite alleles to the SNP hap-

Table 6. Microsatellite Allele Frequencies within The Major Y-Chromosomal Haplotypes *H2* and *H16* in the Finns and Saami^a

Microsatellite locus	Allele	Allele Frequency (%)			
		<i>H2</i>		<i>H16</i>	
		Finns	Saami	Finns	Saami
<i>DYS389-1</i>	9	77	84	1	55
	10	14	6	28	39
	11	9	10	65	4
	12	—	—	5	2
<i>DYS389-2</i>	25	5	3	—	—
	26	64	83	4	2
	27	18	—	30	57
	28	5	—	62	37
	29	5	—	5	6
	30	5	3	—	2
	31	—	3	—	—
32	—	7	—	—	
<i>DYS391</i>	10	95	87	20	19
	11	5	13	80	81
<i>DYS393</i>	12	—	—	1	—
	13	95	98	5	30
	14	5	3	88	70
<i>DYS19</i>	15	—	—	7	—
	14	73	3	98	92
	15	23	57	2	8
	16	5	13	—	—

^aThe most frequent alleles are shown in bold.

lotypes in our population samples. We found that within the major haplotypes *H2* and *H16* in the Finns and Saami, the most frequent microsatellite allele was identical, with the exception of the *DYS19* locus in haplotype *H2* and the *DYS389-1* locus in haplotype *H16* (Table 6). In addition, shared microsatellite haplotypes were observed within the haplotypes formed by the SNPs, but not between them. Therefore, the most common Y-chromosomal haplotypes are shared between the Finns and the Saami, indicating that there have been at least two founding Y-chromosomal lineages in these populations. This finding is in accordance with the archeological data that indicates a dual origin for the Finns, and also with the earlier Y-chromosomal data from the Finns and the Saami (Kittles et al. 1999; Lahermo et al. 1999; Kittles et al. 1999; Semino et al. 2000). However, our data also indicate that within the Finns and the Saami there might be significant subpopulations with regional differences. This is not only seen in the variation of the haplogroup frequencies, but also in the genetic diversity values. For example, the nucleotide diversity in the eastern and northern Finns is strikingly low compared to that of other study populations and the western Finns in particular (Table 5).

METHODS

Population Samples

The Finnish samples originated from two sources. One part of the samples had been collected for a cross-sectional survey on coronary heart disease risk factors in Finland (Vartiainen et al. 1994). The second part of the samples was collected from healthy, unrelated male blood donors (Finnish Red Cross Blood Transfusion Service, Helsinki, Finland). The birthplace of each individual sample donor as well as that of his parents and grandparents were known. The Ob-Ugric and Kola Saami samples were collected by one of the authors (A. Sajantila) during anthropological field surveys to West Siberia in 1995 and the Kola Peninsula in 1996 (see Fig. 1). The collection of the Inari and Skolt Saami and Karelian samples has been described earlier (Lahermo et al. 1999); these were provided by M.-L. Savontaus (University of Turku, Finland). The samples from Norwegian males were provided by M. Stenersen (University of Oslo, Norway). DNA was extracted from the blood samples by standard procedures.

Preparation of Oligonucleotide Arrays

The primer arrays were prepared on microscope glass slides with Teflon lining forming 8 wells of 11 mm in diameter or 12 wells of 5 mm in diameter (Erie Scientific). The slides were prewashed with a 1% Alconox (Aldrich) solution, and rinsed several times with distilled water and ethanol. The minisequencing detection primers with 5'-amino groups (Table 1) were immobilized to isothiocyanate-activated glass surfaces essentially as described by Guo et al. (1994), except that 3-aminopropyltriethoxysilane (Sigma) was used for silanization instead of the methoxy-derivative. The detection primers were dissolved in 400 mM sodium carbonate buffer at pH 9.0 to a final concentration of 25 μ M prior to spotting. Immedi-

ately after spotting, the slides were exposed to vaporized ammonia for 1 h, followed by three washes in distilled water. The arrays were stored at -70°C for up to 8 weeks.

A custom-built, modified industrial robot (Isel) with two TeleChem CPH-2 printing pins controlled by an MCM-310 operating system and NUMO-6.0 software (Merval) was used to print the oligonucleotide detection primers onto the coated slides. For radioactive detection, two adjacent spots 125–150 μ m in diameter of each detection primer at a 400- μ m distance from each other were printed to increase the signal intensities. The distance from the middle of the double spot to the next one was 1000 μ m. The primer order was the same as for the fluorescent detection system, for which the detection primers were applied as spots 125–150 μ m in diameter at a center-to-center distance of 250 μ m (Fig. 2).

PCR Amplification of SNPs

The PCR primers were synthesized by Interactiva Biotechnologie GmbH (Ulm). Their sequences are listed in Table 1. DNA fragments spanning the SNP sites were amplified using a touchdown thermocycling profile in a Programmable Thermal Controller (MJ Tetrad Research). The parameters were: Initial activation of the polymerase at 95°C for 11 min, then 95°C for 30 sec, and $65^{\circ}\text{C} - ^{\circ}\text{C}$ per cycle for 4 min for 5 cycles; 95°C for 30 sec, $60^{\circ}\text{C} - 0.5^{\circ}\text{C}$ per cycle for 2 min, and 68°C for 2 min for 15 cycles; 95°C for 30 sec, 53°C for 30 sec, and 68°C for 2 min for 14 cycles; 68°C for 2 min. One multiplex PCR containing primers for the 18 markers *DYS271* (*M2*), *DYS199* (*M3*), *SRY4064*, *SRY9138*, *SRY10831*, *SRY2627*, *PN3*, *M4*, *M6*, *M7*, *M8*, *M9*, *M11*, *M14*, *M17*, *M19*, *M20*, and *Tat*, and a second multiplex PCR with the 6 markers *PN1*, *PN2*, *M12*, *M13*, *M21*, and *M22* were performed using 10–50 ng of genomic DNA, 3.5 U of AmpliTaq Gold DNA polymerase (Perkin-Elmer), and 200 μ M dNTPs in 100 μ L of DNA polymerase buffer supplied with the enzyme (N808-0244). The primer concentrations ranged from 0.1 to 1.2 μ M, and they had initially been adjusted to yield similar amounts of each DNA fragment and to minimize primer-dimer formation. For occasional reamplification of individual fragments 1 μ L of a 1/100 dilution of the multiplex PCR product was used under the conditions given above. The marker *92R7* was amplified individually under the same conditions as the multiplex PCRs.

SNP Genotyping on Oligonucleotide Arrays

The combined PCR products were precipitated by ethanol, followed by suspension into 80 μ L of 50 mM Tris-HCl at pH 8.5 and 5 mM MgCl_2 buffer containing 0.002 U/ μ L DNase (Ampliscribe T7 Transcription Kit, Epicentre Technologies) and 0.01 U/ μ L shrimp alkaline phosphatase (Boehringer Mannheim, Germany). The mixture was incubated at 37°C for 10 min, followed by inactivation of the enzymes at 95°C for 20 min. Then 20 μ L of a buffer containing 500 mM Tris-HCl at pH 8.0, 250 mM EDTA, 1 M NaCl, and 1% Triton X 100 was added, and the mixture was preheated together with the arrays at 95°C for 1.5 min. Next 10 μ L of this mixture was applied to the arrays to four reaction wells 5 mm in diameter. In case of the larger 11-mm wells, 20 μ L of the mixture was used. The annealing reaction was allowed to proceed in a humid chamber at 37°C for 15 min. The arrays were briefly washed with a solution of 50 mM Tris-HCl at pH 8.0, 25 mM EDTA, 100 mM NaCl, and 0.1% Triton-X 100.

Each minisequencing reaction mixture contained one of the four ^{32}P -labeled ddNTPs (AP Biotech) at a 0.01 μM concentration or TAMRA-labeled ddATP, ddGTP, or ddTTP at a 0.05 μM concentration or ddCTP at a 0.2 μM concentration (NEN Life Science Products), together with the other three unlabeled ddNTPs at a concentration of 0.1 μM in 26 mM Tris-HCl at pH 9.5, 6.5 mM MgCl_2 , and 0.2% Triton X-100 buffer with 0.05 U/ μL DynaSeq DNA polymerase (a kind gift from Finnzymes, Helsinki, Finland) or ThermoSequenase (AP Biotech). The reaction mixture and the arrays carrying the annealed templates were preheated to 68°C for 2 min. The reaction volume for the 5-mm wells was 10 μL , and for the 11-mm wells it was 20 μL . The reaction was allowed to proceed at 68°C for 5 min in a humid chamber, after which the slides were washed three times for 15 min at room temperature in a solution of 90 mM Na-citrate, 900 mM NaCl, and 0.05% *N*-lauroyl-sarcosine. The slides were further washed twice with dH_2O for 15 min, once with 50 mM NaOH for 5 min, and finally with dH_2O for 15 min.

Radioactive signals were detected after an overnight exposure to imaging plates (Fuji, Kanawaga, Japan) using a phosphorimager instrument (Fuji BAS 1500 Bioimaging Analyzer). The signal intensities were measured with the Tina 2.10 software (Raytest). Fluorescence signals were detected using an array scanner (ScanArray 4000, GSI Lumonics), and the signal intensities were measured with the QuantArray analysis software (GSI Lumonics). The ratio between the signal from the reaction for one allele divided by the total signals from the reactions for both alleles was calculated.

Minisequencing in a Microtiter Format

The SNP marker *92R7* was genotyped, and the genotypes of the polymorphic SNP markers *M17*, *SRY10831*, *M12*, *Tat*, and *M9* were confirmed by solid-phase minisequencing in a microtiter plate (Syvänen 1997). For amplification of the individual fragments, the PCR primers were those given in Table 1, except that the 3' PCR primers for markers *92R7* and *SRY10831* were biotinylated in their 5' end. The *92R7* and *SRY10831* sites were amplified individually from 25 ng of genomic DNA using 1.75 U of AmpliTaq Gold DNA polymerase (Perkin-Elmer) with a 0.2 μM PCR primer concentration in a 50- μL volume, except that for the *SRY10831* marker the biotinylated 3' PCR primer was used at a 0.04 μM concentration. The PCR parameters were the same as for the multiplex PCR given above. The markers *Tat*, *M9*, *M17*, and *M12* were amplified individually with both primers at a 0.2 μM concentration under the conditions given above. Of this first amplification product, 1 μL of a 1/100 dilution was subjected to a second amplification with a biotinylated universal primer 5'-GCGGTCCCAAAAGGGTTCAGT to introduce a biotin residue to the PCR products for affinity capture for the minisequencing reaction. The concentration of the biotinylated universal primer was 0.04–0.08 μM , depending on the marker, and it was used together with one of the unbiotinylated primers used in the first amplification at a 0.2 μM concentration under the PCR conditions given above.

The minisequencing primers for the markers *M12* and *M17* were those given in Table 1. For *92R7*, the minisequencing primer was 5'-ATGAACACAAAAGACGTAGAAG; for *SRY10831* it was 5'-GTATCTGACTTTTTCACACAGT; for *M9* it was 5'-GTCTAAATTAAGAAAATAAAGAG; and for *Tat* it was T(15) TGAGTGTAGACTTGTGAATTCA from the complementary DNA strand.

Genotyping of Microsatellite Markers

The PCR primer sequences for the microsatellite markers *DYS19*, *DYS389-I*, *DYS389-II*, *DYS391*, and *DYS393* were obtained from the Genome Database (<http://www.gdb.org/>). The forward primers were fluorescently labeled (DYS 19-TET, DYS 389-FAM, DYS 391-FAM, and DYS 393-TET). The microsatellite markers were amplified using 20 ng of genomic DNA with 1.2 U AmpliTaq Gold DNA polymerase and 200 μM of dNTPs in 100 μL of DNA polymerase buffer supplied with the enzyme (N808-0244). Two separate multiplex PCRs were performed. One of the reactions contained primers for the markers *DYS19*, *DYS389-I*, and *DYS389-II* at 0.3 μM and 0.2 μM concentrations (*DYS389-I* and *DYS389-II* are in one fragment). The cycling parameters were 95°C for 10 min, 94°C for 45 sec for 30 cycles, 55°C for 1 min, 72°C for 1 min, and a final extension at 72°C for 5 min.

The other multiplex PCR mixture contained primers for the markers *DYS391* and *DYS393* at 0.3 μM and 0.1 μM concentrations. The cycling parameters were 95°C for 10 min, 94°C for 1 min for 30 cycles, 51°C for 1 min, and a final extension at 72°C for 5 min.

The PCR products were combined and run in an ABI 377 DNA Sequencer according to the manuals supplied with the sequencer (PE Biosystems). The allelic fragments were detected and sized by the GeneScan 3.1 and ABI-Genotyper 2.0 programs. The fragment sizes were converted into number of repeats using five previously sequenced samples (kindly provided by Manfred Kayser, Leipzig, Germany).

Statistical Analysis

The population genetic analyses of the Y-chromosomal SNP, microsatellite, and haplotype data were performed using the ARLEQUIN (ver 2.0) software package (Schneider et al. 1997). The hierarchic distribution of Y-chromosome diversity was computed using the analysis of molecular variance (AMOVA) software.

ACKNOWLEDGMENTS

We thank Auli Bengs, Minttu Hedman, Kirsti Höök, Liisa Kauppi, Minna Levander, and Päivi Tainola for assistance with laboratory work, and Paavo Niini for technical expertise with the arrayer. We are grateful to Michael Hammer, Chris Tyler-Smith, and Peter Underhill for unpublished sequences, and to Manfred Kayser for the control microsatellite sequences. We thank Finnzymes for providing the DynaSeq DNA polymerase free of charge. This work was funded by EC Biomed2, Contract no. BMH4-CT97-2013, the Instrumentarium Foundation, and the Academy of Finland (project no. 42183).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altheide, T.K. and Hammer, M.F. 1997. Evidence for possible Asian origin of YAP+ Y chromosomes. *Am. J. Hum. Genet.* **61**: 462–466.
- Barbujani, G., Magagni, A., Minch, E., and Cavalli-Sforza, L.L. 1997. An apportionment of human DNA diversity. *Proc. Natl. Acad. Sci.* **94**: 4516–4519.

- Bianchi, N.O., Bailliet, G., Bravi, C.M., Carnese, R.F., Rothhammer, F., Martinez-Marignac, V.L., and Pena, S.D. 1997. Origin of Amerindian Y-chromosomes as inferred by the analysis of six polymorphic markers. *Am. J. Phys. Anthropol.* **102**: 79–89.
- Di Rienzo, A. and Wilson, A.C. 1991. Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci.* **88**: 1597–1601.
- Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., and Mattick, J.S. 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. *Nucleic Acids Res.* **19**: 4008.
- Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R., and Smith, L.M. 1994. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **22**: 5456–5465.
- Hacia, J.G. 1999. Resequencing and mutational analysis using oligonucleotide microarrays. *Nat. Genet.* **21**: 42–47.
- Hacia, J.G., Sun, B., Hunt, N., Edgemon, K., Mosbrook, D., Robbins, C., Fodor, S.P.A., Tagle, D., and Collins, F.S. 1998. Strategies for mutational analysis of the large multiexon ATM gene using high-density oligonucleotide arrays. *Genome Res.* **8**: 1245–1258.
- Hammer, M.F. 1995. A recent common ancestry for human Y chromosomes. *Nature* **378**: 376–378.
- Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., et al. 1997. The geographic distribution of human Y chromosome variation. *Genetics* **145**: 787–805.
- Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A.R., and Zegura, S.L. 1998. Out of Africa and back again: Nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**: 427–441.
- Hurles, M.E., Veitia, R., Arroyo, E., Armerentos, M., Bertranpetit, J., Pérez-Lezaun, A., Bosch, E., Shlumukova, M., Cambon-Thomsen, A., McElreavey, K., et al. 1999. Recent male-mediated gene flow over a linguistic barrier in Iberia, suggested by analysis of a Y-chromosomal DNA polymorphism. *Am. J. Hum. Genet.* **65**: 1437–1448.
- Jobling, M.A., Pandya, A., and Tyler-Smith, C. 1997. The Y chromosome in forensic analysis and paternity testing. *Int. J. Leg. Med.* **110**: 118–124.
- Karafet, T., Zegura, S.L., Vuturo-Brady, J., Posukh, O., Osipova, L., Weibe, V., Romero, F., Long, L.C., Harihara, S., Jin, F., et al. 1997. Y chromosome markers and trans-Bering Strait dispersals. *Am. J. Phys. Anthropol.* **102**: 301–314.
- Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Kniff, P., et al. 1999. Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**: 817–831.
- Kittles, R.A., Perola, M., Peltonen, L., Bergen, A.W., Aragon, R.A., Virkkunen, M., Linnoila, M., Goldman, D., and Long, J.C. 1998. Dual origins of Finns revealed by Y chromosome haplotype variation. *Am. J. Hum. Genet.* **62**: 1171–1179.
- Kittles, R.A., Bergen, A.W., Urbanek, M., Virkkunen, M., Linnoila, M., Goldman, D., and Long, J.C. 1999. Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: Evidence for a male-specific bottleneck. *Am. J. Phys. Anthropol.* **108**: 381–399.
- Kwok, W., Tyler-Smith, C., Mendonca, B.B., Hughes, I., Berkovitz, G.D., Goodfellow, P.N., and Hawkins, J.R. 1996. Mutation analysis of the 2 kb 5' to SRY in XY females and XY intersex subjects. *J. Med. Genet.* **33**: 465–468.
- Lahermo, P., Savontaus, M.L., Sistonen, P., Beres, J., de Knijff, P., Aula, P., and Sajantila, A. 1999. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. *Eur. J. Hum. Genet.* **7**: 447–458.
- Lell, J.T., Brown, M.D., Schurr, T.G., Sukernik, R.I., Starikovskaya, Y.B., Torroni, A., Moore, L.G., Troup, G.M., and Wallace, D.C. 1997. Y chromosome polymorphisms in native American and Siberian populations: Identification of native American Y chromosome haplotypes. *Hum. Genet.* **100**: 536–543.
- Mathias, N., Bayes, M., and Tyler-Smith, C. 1994. Highly informative compound haplotypes for the human Y chromosome. *Hum. Mol. Genet.* **3**: 115–123.
- Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L., and Syvänen, A.-C. 1997. Minisequencing: A specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* **7**: 606–614.
- Pastinen, T., Perola, M., Niini, P., Terwilliger, J., Salomaa, V., Vartiainen, E., Peltonen, L., and Syvänen, A.-C. 1998. Array-based multiplex analysis of candidate genes reveals two independent and additive genetic risk factors for myocardial infarction in the Finnish population. *Hum. Mol. Genet.* **7**: 1453–1462.
- Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., and Syvänen, A.-C. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* **10**: 1031–1042.
- Ruiz-Linares, A., Ortiz-Barrientos, D., Figueroa, M., Mesa, N., Munera, J.G., Bedoya, G., Velez, I.D., Garcia, L.F., Perez-Lezaun, A., Bertranpetit, J., et al. 1999. Microsatellites provide evidence for Y chromosome diversity among the founders of the New World. *Proc. Natl. Acad. Sci. USA* **96**: 6312–6317.
- Sajantila, A. and Pääbo, S. 1995. Language replacement in Scandinavia. *Nat. Genet.* **11**: 359–360.
- Sajantila, A., Lahermo, P., Anttinen, T., Lukka, M., Sistonen, P., Savontaus, M.L., Aula, P., Beckman, L., Tranebjaerg, T., Gedde-Dahl, T., et al. 1995. Genes and languages in Europe: An analysis of mitochondrial lineages. *Genome Res.* **5**: 42–52.
- Schneider, S., Kueffer, J.-M., Roessli, D., and Excoffier, L. 1997. Arlequin ver 1.1: A software for population genetic data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.
- Seielstad, M.T., Hebert, J.M., Lin, A.A., Underhill, P.A., Ibrahim, M., Vollrath, D., and Cavalli-Sforza, L.L. 1994. Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition. *Hum. Mol. Genet.* **3**: 2159–2161.
- Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., De Benedictis, G., Francalacci, P., Kouvatsi, A., Limborska, S., et al. 2000. The genetic legacy of paleolithic *Homo sapiens sapiens* in extant Europeans: A Y chromosome perspective. *Science* **290**: 1155–1159.
- Shen, P., Wang, F., Underhill, P.A., Franco, C., Yang, W.-H., Roxas, A., Sung, R., Lin, A.A., Hyman, R.W., Vollrath, D., et al. 2000. Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl. Acad. Sci.* **97**: 7354–7359.
- Shuber, A.P., Grondin, V.J., and Klinger, K.W. 1995. A simplified procedure for developing multiplex PCRs. *Genome Res.* **5**: 488–493.
- Sigurðardóttir, S., Helgason, A., Gulcher, J.R., Stefansson, K., and Donnelly, P. 2000. The mutation rate in the human mtDNA control region. *Am. J. Hum. Genet.* **66**: 1599–1609.
- Syvänen, A.-C. 1997. Solid-phase minisequencing. In *Laboratory methods for the detection of mutations and polymorphisms in DNA* (ed. G.R. Miller), pp. 53–64. CRC Press, Florida.
- Syvänen, A.-C. 1999. From gels to chips: "Minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* **13**: 1–10.
- Syvänen, A.-C., Aalto-Setälä, K., Harju, L., Kontula, K., and Söderlund, H. 1990. A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. *Genomics* **8**: 684–692.
- Thomson, R., Pritchard, J.K., Shen, P., Oefner, P.J., and Feldman, M.W. 2000. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl. Acad. Sci.* **97**: 7360–7365.
- Tyler-Smith, C. 1999. Y-chromosomal DNA markers. In *Genomic diversity: Applications in human population genetics* (eds. S.S. Papiha et al.), pp. 65–73. Plenum, New York.
- Underhill, P.A., Jin, L., Zemans, R., Oefner, P.J., and Cavalli-Sforza, L.L. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci.* **93**: 196–200.
- Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., and Oefner, P.J. 1997.

- Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* **7**: 996–1005.
- Vartiainen, E., Puska, P., Jousilahti, P., Korhonen, H.J., Tuomilehto, J., and Nissinen, A. 1994. Twenty-year trends in coronary risk factors in North Karelia and in other areas of Finland. *Int. J. Epidemiol.* **23**: 495–504.
- Veitia, R., Ion, A., Barbaux, S., Jobling, M.A., Souleyreau, N., Ennis, K., Ostrer, H., Tosi, M., Meo, T., Chibani, J., et al. 1997. Mutations and sequence variants in the testis-determining region of the Y chromosome in individuals with a 46,XY female phenotype. *Hum. Genet.* **99**: 648–652.
- Whitfield, L.S., Sulston, J.E., and Goodfellow, P.N. 1995. Sequence variation of the human Y chromosome. *Nature* **378**: 379–380.
- Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhover, W., Fretwell, N., Jobling, M.A., Harihara, S., et al. 1997. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**: 1174–1183.

Received July 18, 2000; accepted in revised form December 29, 2000.