Int J Legal Med (2008) 122:449–456
DOI 10.1007/s00414-008-0250-6

ORIGINAL ARTICLE

The effect of number of loci on geographical structuring and forensic applicability of Y-STR data in Finland

Jukka U. Palo · Markus Pirttimaa · Auli Bengs · Vivian Johnsson · Ismo Ulmanen · Matti Lukka · Bjarne Udd · Antti Sajantila

Received: 5 December 2007 / Accepted: 17 April 2008 / Published online: 12 June 2008

Abstract The Y-chromosomal diversity among Finnish males is characterized by low diversity and substantial geographical substructuring. In a 12-locus data set (PowerPlexY), especially the eastern parts of the country showed low levels of variation, and the western, middle, and eastern parts of Finland differed from each other by their Y-short tandem repeat (STR) haplotype frequencies (Palo et al., Forensic Sci Int Genet 1:120–124, 2007). In this paper, we have analyzed geographical patterns of Y-STR diversity using both 12-locus (PowerPlexY) and 17-locus (Yfiler) data sets from the same set of geographically structured samples. In the larger data set, the haplotype diversity is significantly higher, as expected. The geographical distribution of haplotypes is similar in both data sets, but the level of interregional differences is significantly lower in the Yfiler data. The implications of these observations on the forensic casework are discussed.

J. U. Palo (✉) · A. Sajantila
Department of Forensic Medicine, University of Helsinki,
PO BOX 40, 00014 Helsinki, Finland
e-mail: jukka.palo@helsinki.fi

M. Pirttimaa · A. Bengs · V. Johnsson
Forensic Laboratory, National Bureau of Investigation,
PO BOX 285, 01301 Vantaa, Finland

I. Ulmanen · M. Lukka
Paternity Testing Laboratory, National Public Health Institute,
Mannerheimintie 166,
00300 Helsinki, Finland

B. Udd
Vaasa Central Hospital,
Hietalahdenkatu 2-4,
65130 Vaasa, Finland

Keywords Finland · Y-STR · AmpF*l*STR Yfiler · PowerPlexY · Population subdivision

Introduction

During the past decade, the analyses of Y-chromosomal markers have given valuable insights into human evolution and phylogeography [7, 10, 26, 27]. These markers reside in the male-specific, nonrecombining portion of the genome and thus allow the reconstruction of male lineages. The Y-chromosomal markers, mainly Y-specific short tandem repeats (STRs), are widely used as forensic markers as well, being useful, e.g., in rape cases to identify a male perpetrator from a mixture of female and male DNA. However, for the forensic community, the Y-chromosomal properties useful for evolutionary studies can be problematic. In addition to the fact that Y-chromosomal markers are individualizing only on the level of male lineages (i.e., men of the same male descent lineage carry identical haplotypes), the high level of drift at Y-chromosome can introduce problems for statistical interpretations. First, low Y-STR diversity in small and isolated populations unavoidably denotes reduced discriminatory power. Second, the presence of substructure in the database may lead to erroneous frequency estimates for the Y-STR profiles obtained. As the product rule does not apply for these linked markers, the frequencies of the profiles can only be estimated from homogeneous, representative databases [9] (see also [28]).

The Finnish population history entails founder effects, as well as subsequent bottlenecks and local isolation [16]. Consequently, the Finns harbor significantly lower Y-STR diversity than observed in other European populations [8, 13, 23, 24]. The reduction of variation is more drastic in the

eastern parts of the country, which also partly contributes to the geographical substructure described in the Finnish Y-STR variation [8, 14, 17]. Recently, Palo et al. [17] showed that in the 12-locus Y-STR data the western, middle, and eastern parts of Finland form separate clusters, with similar levels of differentiation as observed among major European population groups. The haplotypic differentiation relevant in forensic casework appears moderate, yet it may have significance in the case of certain haplotypes. In these cases, using a pooled data set would lead up to sevenfold underestimation or overestimation of the regional haplotype frequency.

The most limiting factor for the effective use of Y-STR markers in Finland is the low Y-chromosomal diversity observed—a problem, which could potentially be solved by including more loci. Several multiplex kits are commercially available for typing Y-STR loci (see e.g. [15]). The widely used PowerPlexY (Promega) and AmpF*l*STR Yfiler (Applied Biosystems) include the loci of the "European minimal haplotype" [11] and three or eight additional loci, respectively.

Here, we have evaluated the magnitude of differences in diversity and geographical distribution between 17-locus Yfiler and a 12-locus subset corresponding to PowerPlexY haplotypes in a Finnish population sample. The main aim was to characterize how the number of Y-STR loci affects the level and geographic distribution of Y-chromosomal diversity.

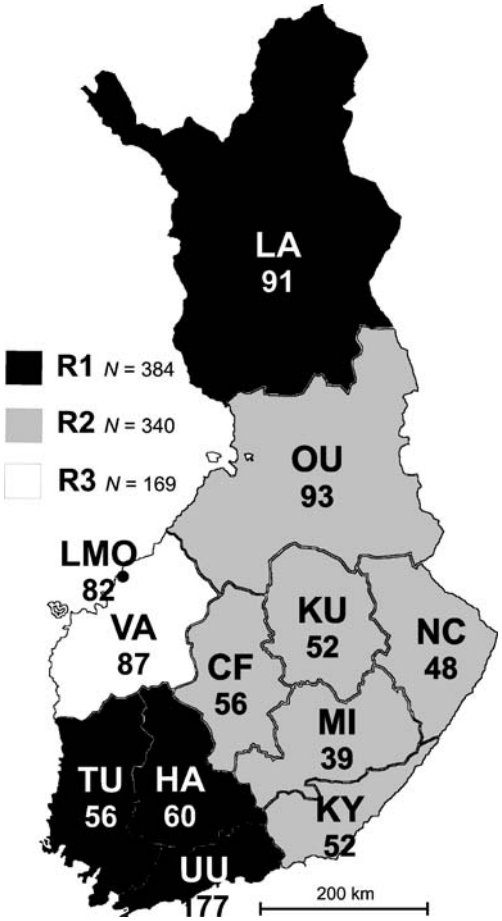

Fig. 1 The locations of the sampled subpopulations in Finland. The *numbers* indicate the sample sizes, and *colors* denote the clustering of populations based on Yfiler F_{ST} distances. For abbreviations see text

Materials and methods

Altogether, 893 unrelated Finnish males were sampled for this study, with 605 samples included in [17]. The remaining samples were collected with informed consents and have not been previously genotyped for Y-STR markers. The samples were assigned to 12 subpopulations according to the current place of residence of the donors: Turku (TU), Uusimaa (UU), Häme (HA), Vaasa (VA), Kymi (KY), Central Finland (CF), Mikkeli (MI), Kuopio (KU), Northern Carelia (NC), Oulu (OU), Lapland (LA) and Larsmo (LMO; Fig. 1). These subpopulations correspond to the former Finnish administrational provinces, except LMO which is a relatively isolated island region of the Vaasa province. This locality was included in the study as it is an almost exclusively Swedish-speaking community. Currently, Finland has a ca. 6% Swedish-speaking minority inhabiting mainly the coastal regions since the early medieval period.

DNA was extracted using standard methods, and the concentration of the Y-chromosomal DNA in each sample was quantified using Quantifiler™ Y Human Male DNA Quantification Kit (Applied Biosystems). All samples were genotyped for 17 Y chromosomal STR loci (DYS456, DYS389I, DYS390, DYS389II, DYS458, DYS19, DYS385a/b, DYS393, DYS391, DYS439, DYS635, DYS392, Y_GATA_H4, DYS437, DYS438, DYS448; this order of the loci is maintained throughout the article) using the AmpF*l*STR Yfiler genotyping kit and 1.0 ng of template DNA. Amplification products were resolved on ABI Prism ®3100 Genetic Analyzer (Applied Biosystems) automated sequencer and analyzed using GeneScan v. 3.7 and Genotyper v. 3.7 NT software (Applied Biosystems). Pooled (non-regional) 11-locus haplotype data is available in the Y-chromosome Haplotype Reference Database (www.yhrd.org).

From this complete Yfiler data set, the PowerPlexY-corresponding data set was formed by removing the loci DYS456, DYS458, DYS635, Y_GATA_H4, and DYS448. The data sets are hereafter referred to as Yfiler and PPY, respectively. For both analyses, the repeat number of DYS389I was subtracted from that of DYS389II.

The statistical analyses were performed with both sets using the software ARLEQUIN 3.1 [4] unless otherwise indicated and tested for statistical significance using 10,000 randomizations. Y-STR diversity was examined by calculating the number of haplotypes (*A*) and by estimating the

haplotype ($\hat{H}$) diversity. Allelic richness (*AR*) was estimated for each population assuming a sample size of the smallest population (MI, N=39) for all the subpopulations using software contrib 1.02 [19]. Genetic differentiation between the subpopulations was assessed by conventional F_{ST} estimates based on haplotype frequencies, as well as by taking the distances between haplotypes into account (Φ_{ST}). Negative distances were considered as zero. The F_{ST} and Φ_{ST} interpopulation distances in the Yfiler and PowerPlexY data were visualized by neighbor-joining (NJ) trees constructed with MEGA ver. 4 [25]. A clustering minimizing the within group and maximizing the among-group variation was manually searched for by hierarchical grouping of the subpopulations (into regions; see [23]) based on the F_{ST} estimates from Yfiler data. The levels of among-region (F_{CT}) and within-region (F_{SC}) haplotypic variation were estimated using AMOVA.

As incomplete Y-STR profiles are regularly obtained from trace evidence material in forensic casework; we also estimated the levels of geographical subdivision of partial Yfiler profiles. For this assessment, two approaches were used. First, each locus was separately removed from the data set, yielding 16-locus data sets (N=17). The F_{CT} values in these data sets were then compared with values estimated from full profiles assuming the geographical structure described for the Yfiler data. Secondly, the amplification success of each locus was assessed from partial profiles (N=44) obtained in casework at the National Bureau of Investigation Forensic Laboratory. From the data, one to seven loci were sequentially removed in the order of probability of non-amplification, and the changes in the variance components were recorded.

Results

The PPY data set analyzed here was created from the Yfiler data by removing the five Yfiler-specific loci from the data. This was considered to represent actual PowerPlexY genotyping well, as good concordance between the Yfiler and PowerPlexY results has been reported [5]. In addition, 605 samples included in this study have been genotyped previously using the PowerPlexY-kit [17]. No discrepancies were found in this sample subset between the two kits.

Diversity

As expected, the Yfiler data revealed significantly more variation than the PPY. Altogether 517 unique haplotypes (*A*) and $\hat{H}$=0.992±0.001 were observed in the Yfiler data set, whereas these values were *A*=319 and $\hat{H}$=0.965±0.004 in the PPY (Table 1). The numbers of singletons for Yfiler were 412 (79.7% of the haplotypes) and for PPY, 236 (74.0%). The most common haplotype in the PPY data was observed 141 times, with a frequency p>0.20 in four subpopulations (p=0.256 in MI). With the addition of five loci, this haplotype was dissected into 21 unique haplotypes of which 9 were observed once and the most common, 67 times. This Yfiler haplotype had the highest frequency of p=0.135 in the subpopulation KY.

There were statistically significant differences in the PPY haplotype diversity between the subpopulations, the values ranging from $\hat{H}$=0.933±0.028 in KY to $\hat{H}$=0.986±0.007 in TU. The variation of the haplotype diversities among subpopulations was notably higher than in the Yfiler data (Table 1).

Table 1 Basic statistics resolved with the two Y-STR genotyping kits

Samples		*N*	Yfiler			PowerPlexY		
			A	*AR*	$\hat{H}$±SD	*A*	*AR*	$\hat{H}$±SD
Häme	HA	60	54	35.3	0.996±0.004	43	28.9	0.972±0.013
Central Finland	CF	56	47	33.0	0.990±0.007	30	22.1	0.946±0.019
Uusimaa	UU	177	140	34.3	0.991±0.003	102	28.3	0.966±0.009
Vaasa	VA	87	71	33.9	0.993±0.004	57	28.3	0.971±0.011
North Carelia	NC	48	40	32.0	0.983±0.011	32	25.8	0.936±0.029
Mikkeli	MI	39	32	31.0	0.987±0.010	22	21.0	0.908±0.034
Turku	TU	56	56	38.0	1.000±0.003	42	30.4	0.986±0.007
Kymi	KY	52	43	32.1	0.982±0.012	33	24.9	0.933±0.028
Oulu	OU	93	65	30.8	0.983±0.007	53	26.7	0.957±0.015
Lappi	LA	91	71	33.4	0.992±0.003	53	26.5	0.966±0.010
Larsmo	LMO	82	56	29.9	0.984±0.005	44	24.5	0.966±0.008
Kuopio	KU	52	42	31.4	0.985±0.009	31	27.8	0.943±0.022
ALL		893	517	35.1	0.992±0.001	319	26.1	0.965±0.004

N Sample size, *A* number of haplotypes, *AR* allelic richness, the probable number of alleles observed among 39 individuals, $\hat{H}$ haplotype diversity

Population subdivision

On average, the haplotypic differentiation, as measured by F_{ST}, was an order of magnitude lower than the Φ_{ST} in both data sets.

In the Yfiler data, the interpopulation F_{ST} distances varied from 0 to 0.010 (LMO–KY), with 22 out of 66 comparisons being significantly larger than zero. The subpopulation LMO differed significantly from all the other populations. In contrast, the Φ_{ST} values ranged from Φ_{ST}=0 to 0.227 with the largest value observed again between LMO and KY. Over half (36) of the Φ_{ST} distances were significantly larger than zero in the Yfiler data set.

In the PPY data, the differentiation between subpopulations was notably higher than with the 17-locus Yfiler. The maximum F_{ST}=0.037 was observed between LMO and MI. In contrast, the Φ_{ST} values estimated from the PPY data were very similar to those observed in the Yfiler-data set, with the maximum Φ_{ST}=0.218 observed between subpopulations LMO and KY. On the whole, the F_{ST} distances differed more between the data sets than the Φ_{ST} distances, yielding a somewhat different NJ-tree topology in the F_{ST}, but not in the Φ_{ST} distances (Fig. 2).

The Yfiler data can be clustered into four groups (regions) in the hierarchical clustering. The regions consist of one to six subpopulations *R1*: TU, UU, HA, LA; *R2*: KY, MI, CF, KU, NC, OU; *R3*: VA, and *R4*: LMO (Fig. 1). For this grouping, the AMOVA gives statistically significant among-region variation F_{CT}=0.38% (P=0.0003) and non-significant F_{SC}=0. Although LMO is differentiated from all the other populations, as part of the VA province region, R4 was included in R3 for the subsequent analyses. This increased the within-regions and reduced the among-regions variance components, but there was no qualitative change (F_{CT}=0.22%, P=0.0008 and F_{SC}=0.09%, P=0.0784). Assuming this structure, the PPY data gives F_{CT}=0.67% (P=0.0012) and F_{SC}=0.3% (P=0.0002).

As the geographical distribution of haplotypes differed between the Yfiler and PPY data, we also examined this effect with partial profiles consisting of 10–16 loci. The analysis revealed differences in the obtained among-region differentiation, depending on the missing locus. The absence of loci DYS458, DYS635, or DYS439 marked an increase of >50% in the F_{CT} values, whereas the change was negligible in case of eight of the loci (Fig. 3a). In fact, the absence of DYS438 or DYS390 lowered the among-region differentiation.

The order of removal of multiple loci was based on dropout frequencies observed during actual casework. Among these 44 cases, the non-amplified loci were DYS389II (84% of the cases), DYS19 (68%), DYS439 (68%), DYS390 (45%), DYS392 (43%), and DYS385a/b (41%). The rest of the loci were unsuccessfully amplified in less than 20% of the cases, with DYS393 and DYS391 successful in all. The absence of the most commonly unamplified loci had little impact on the geographical differentiation before three or more loci are missing (Fig. 3b). The within-region variation remained roughly the same despite reductions in the amount of data.

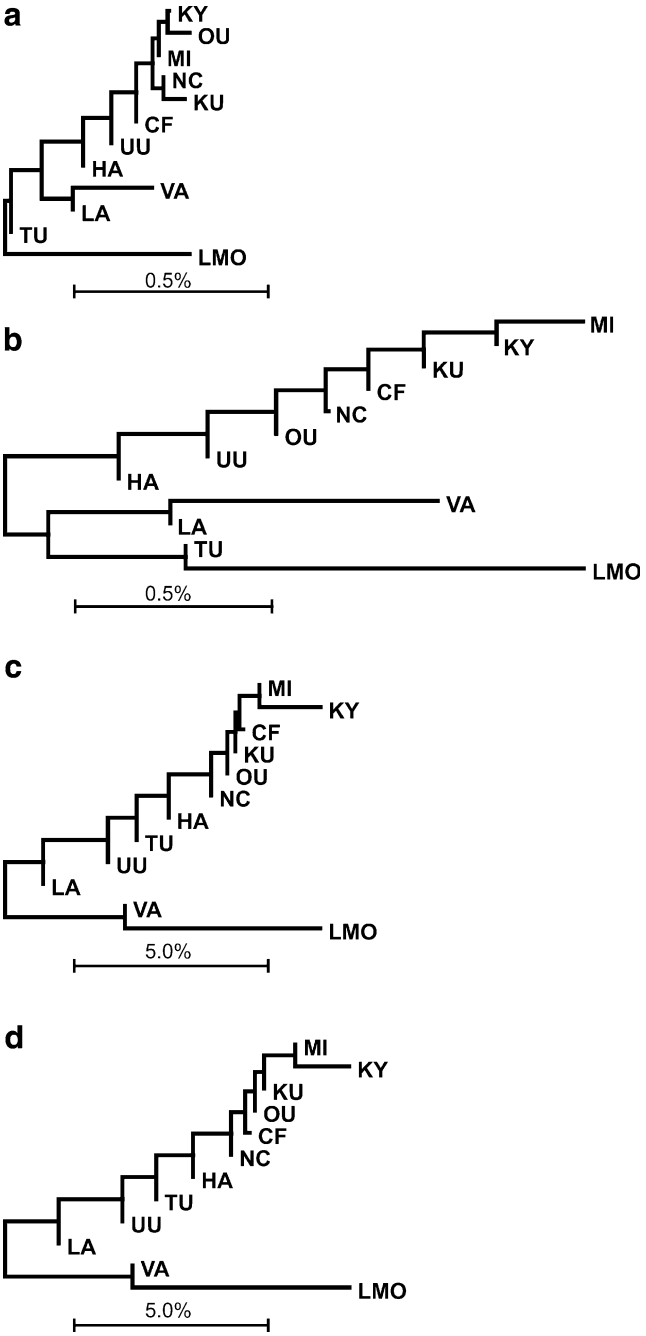

Fig. 2 Midpoint-rooted NJ-trees from two different distance indices and data sets. **a** F_{ST} Yfiler; **b** F_{ST} PowerPlexY; **c** Φ_{ST} Yfiler; **d** Φ_{ST} PowerPlexY. Note the different scale in the F_{ST} and Φ_{ST} trees

Discussion

Previous Y-chromosomal studies have shown that the Finnish Y-chromosomal gene pool is characterized by low diversity and

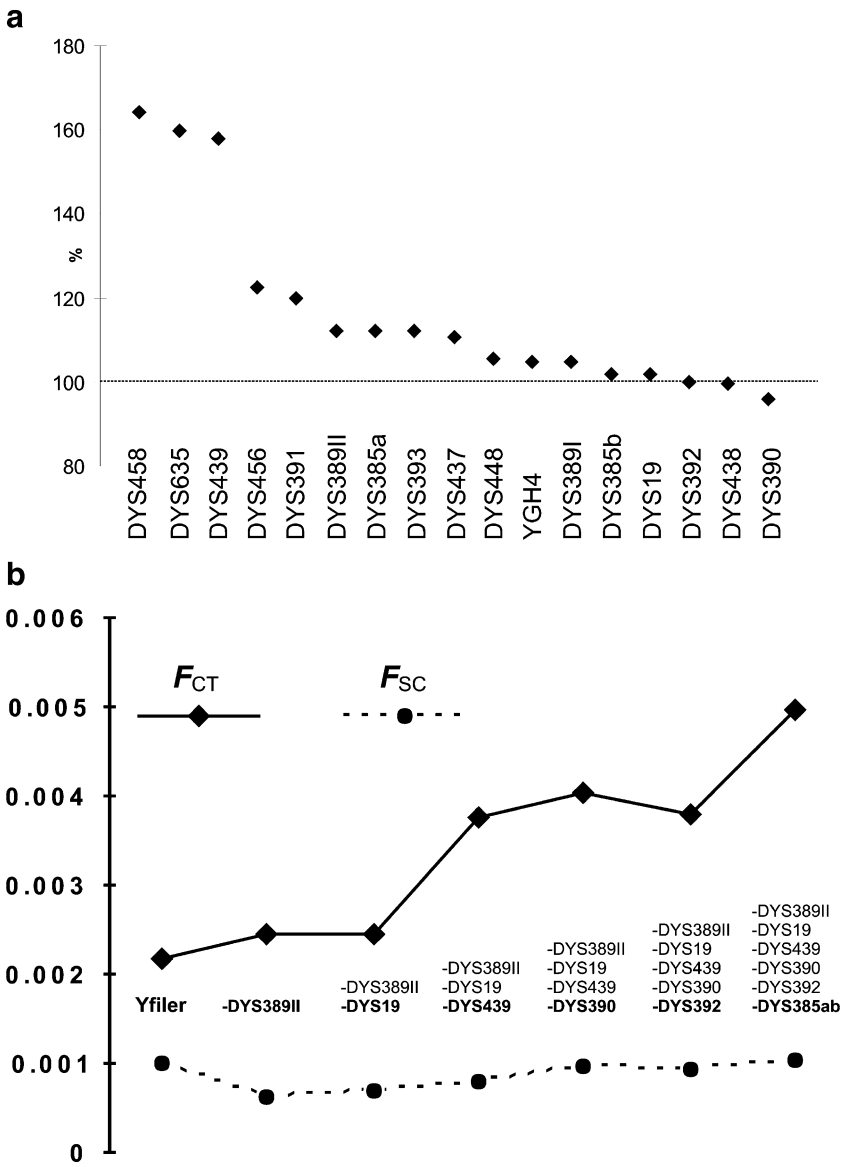

Fig. 3 The effect of locus removal on the differentiation indices, assuming the geographical clustering (see Fig. 1). The absent locus is indicated on the *x*-axis. **a** Relative increase of F_{CT} values observed in 16-locus partial profiles compared to full Yfiler data. Note that 385a and b were considered as separate loci here. **b** F_{CT} (*solid line*) and F_{SC} (*hatched line*) values obtained after removing one to ten loci in the order of observed non-amplification probability

substantial geographical structuring [8, 14, 17]. Both of these factors signify a challenge for forensic application of Y-chromosomal markers in Finland. However, the number of markers, e.g., between the 12-locus PowerPlexY and 17-locus Yfiler amplification kits, can have a significant effect on the magnitude of these problems. Increasing the amount of scored loci from 12 to 17 loci expectedly increased the discriminatory power of the Y-STRs but also reduced the interregional differentiation, both desirable attributes in forensic casework. The comparison between the two data sets in the presence of population substructure revealed several noteworthy aspects with relevance for the forensic work.

Diversity

The low diversity in the PPY data is obvious: for instance, the four most common PPY haplotypes were encountered in 254 individuals (28.4%). These haplotypes segregated into 62 unique haplotypes with the Yfiler data. Compared to the PPY, the additional five loci included in the Yfiler kit increased the number of unique haplotypes by 62% in the total data. Although there was also a significant increase in the haplotype diversity (i.e. power of discrimination; Table 1), one Finnish subpopulation showed point estimates as low as $\hat{H}=0.982$ in the Yfiler data set. The overall $\hat{H}=0.992\pm0.001$ is lower than reported in other populations as well. For the sake of comparison, analyses of 250 males from Portugal [1] and 500 males from Brazil [18] have recently reported Yfiler haplotype diversities $\hat{H}>0.999$. The highest absolute haplotype frequency in both of these studies was 3, corresponding to 1.2% and 0.6% in the Portuguese and Brazilian samples, respectively. In contrast, among the 897 Finnish males the most common Yfiler haplotype was observed 67 times, i.e., in 7.5% of the males

studied with relative frequencies almost twice as high in some subpopulations.

Population subdivision

Population substructure presents another challenge for forensic geneticists: the reliable estimation of haplotype frequencies must be based on homogeneous database(s), or alternatively, the observed frequencies must be appropriately corrected based on subdivision information (e.g., [3], see also [2])

The geographical substructure among the Finnish males was notable when measured with the Φ_{ST} values, reaching values as high as Φ_{ST}=0.227 in the Yfiler data. This is rather extreme, given that, e.g., subpopulations Larsmo and Kymi are separated by mere 400 km, with no apparent physical dispersal barriers between them. In contrast, the deepest divergence between the European populations included in [23] is $\Phi_{ST}\approx$0.08 (minimal haplotype data), and several studies including U.S. populations have described significant differences mainly between different ethnic groups (e.g., European and African Americans; [3, 12, 20]. Nonsignificant differentiation is recently also reported between geographically relatively widely separated populations in Russia [21]. From the forensic point of view, the substructure effects in Finland are exacerbated by the fact that clearly distinct Y-STR haplotypes dominate in different subpopulations (Fig. 4). This may be explained by gene flow from Scandinavia that occurred after the initial southeastern colonization of Finland and extended only to the southwestern parts of Finland. The eastern haplogroup N3 is the most common in Finnish males and is common in all parts of the country, whereas western Finland also harbor significant proportions of the Scandinavian haplogroup I1a [14]. Indeed, the most common Y-STR haplotype 14-14-24-16-17-14-11.13-14-11-10-21-14-12-14-10-19 was confirmed to be associated with haplogroup N3 by SNP typing (data not shown). This haplotype is the most common haplotype in other subpopulations except in VA and LMO, where 14-12-23-14-16-14-14.14-13-10-10-22-11-11-16-10-20 and 14-13-23-16-15-14-14.14-13-10-10-22-11-11-16-10-20 had the highest frequencies, respectively. These haplotypes, associated with haplogroup I, were not encountered in most of the other subpopulations in the current sample.

The difference in haplogroup distribution in Finland is manifested as an order of magnitude higher Φ_{ST} than F_{ST} values. Interestingly, the number of loci had notable effect on the F_{ST} but not on the Φ_{ST} differentiation. The addition of data reduced the observed F_{ST} distances to roughly one third of those obtained with the PPY. The number of loci affected also the clustering of populations as seen in the NJ-tree topologies (Fig. 2). The haplotypic differentiation is the most relevant measure in forensics, as deductions are based on the identity or nonidentity of the profiles obtained from the sample material and the reference sample. This observation suggests again that increasing the amount of loci would be a sound approach in forensics, but as of yet, it is unclear how population-specific these effects are.

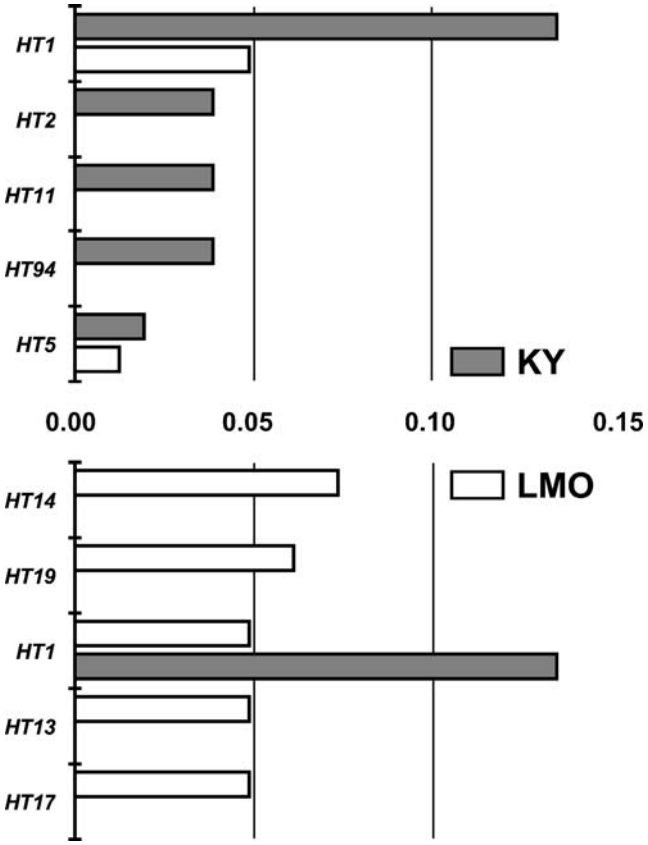

Fig. 4 The frequencies of the five most common haplotypes observed in subpopulations KY (*upper graph*) and LMO (*lower graph*). The frequency of each haplotype is plotted in both subpopulations. *Haplotype names* refer to their frequency rankings in Finland

It is important to note that in the presence of substructure, partial Y-STR profiles may have geographical distribution patterns which differ from that of the full profiles. The magnitude of this change depends closely on the locus/loci missing with the effect on the among-region variation correlating with the variability at the missing locus. The absence of certain loci, especially DYS458, DYS635, or DYS439 raised the among-region haplotypic variation from 0.22% up to 0.36%. These loci had relatively uniform allele frequencies among the regions. However, the loci observed to have low amplification success in real forensic samples have less effect on the levels of genetic structure.

Conclusions

Considering forensic casework in Finland, the Yfiler kit outperforms PowerPlexY because of its significantly higher

discrimination power, but even with this additional data, the power remains low in absolute terms. The low discrimination power is probably the most compromising factor in the application of Y-STR data in Finland.

Finland also shows rather extreme levels of Y-chromosomal differences between different regions. The significant differentiation between subpopulations suggests that using a single database to estimate profile frequencies can introduce a bias for assessing the power of evidence in forensic casework. The most common haplotypes in different regions differ from each other, often belonging to different haplogroups. Despite this rather drastic difference (see Fig. 4), the differentiation measured by F_{ST} remains moderate at most, suggesting that, in certain cases, the summarizing *F*-statistics may not be the best possible barometer of substructure effects in the forensic context.

These compromising factors were observed in both the 12-locus PPY and the 17-locus Yfiler data sets. However, there were significant differences in the diversity patterns revealed by these two amplification kits. Scoring more loci increased the Y-STR diversity and also lowered the level of interregional differentiation, which both stem partly from the increased amount of singletons in the larger data set. Even though the interregional haplotypic differentiation in Finland appears low when measured with summary statistics such as AMOVA, individual haplotypes may have widely differing frequencies. The statistically significant differences between regions advocate the use of separate databases for the estimation of Y-STR profile frequencies in forensic casework. The forensic community has not yet issued any unambiguous guidelines on how the population substructure should be taken into account in statistical interpretations of the matches (but see [6, 22]). At present, it is of vital importance to recognize the potential population substructure patterns; in the case of haploid markers, this calls for genotyping of representative, geographically structured sample collections.

Acknowledgements The authors wish to thank Ms. Marika Merisalo and Mrs. Airi Sinkko for technical assistance, and four anonymous referees for their comments on the manuscript. Financial support was provided by the Academy of Finland: grants 1109265 and 1111713 to JUP; grant 80578 to AS within the framework of the European Science Foundation EUROCORES programme "The Origin of Man, Language and Languages".

References

1. Alves C, Gomes V, Prata MJ, Amorim A, Gusmão L (2007) Population data for Y-chromosome haplotypes defined by 17 STRs (AmpFISTR YFiler) in Portugal. Forensic Sci Int 171:250–255
2. Balding DJ, Nichols RA (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. Forensic Sci Int 64:125–140
3. Budowle B, Adamowicz M, Aranda XG et al (2005) Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. Forensic Sci Int 150:1–15
4. Excoffier L, Laval LG, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetic data analysis. Evol Bioinf Online 1:47–50
5. Gross AM, Berdos P, Ballantyne J (2006) Y-STR concordance study between Y-Plex (TM) 5, Y-Plex (TM) 6, Y-Plex (TM) 12, Powerplex (R) Y, Y-Filer (TM), MPI, and MPII. J Forensic Sci 51:1423–1428
6. Gusmão L, Butler JM, Carracedo A et al (2006) DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Int J Legal Med 120:191–200
7. Hammer MF, Zegura SL (2002) The human Y chromosome haplogroup tree: nomenclature and phylogeography of its major divisions. Ann Rev Anthropol 31:303–321
8. Hedman M, Pimenoff V, Lukka A, Sistonen P, Sajantila A (2004) Analysis of 16 Y STR loci in the Finnish population reveals a local reduction in the diversity of male lineages. Forensic Sci Int 142:37–43
9. Jobling MA, Pandya A, TylerSmith C (1997) The Y chromosome in forensic analysis and paternity testing. Int J Legal Med 110:118–124
10. Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612
11. Kayser M, Caglia A, Corach D et al (1997) Evaluation of Y-chromosomal STRs: a multicenter study. Int J Legal Med 110:125–133 141–129
12. Kayser M, Brauer S, Schadlich H et al (2003) Y chromosome STR haplotypes and the genetic structure of US populations of African, European, and Hispanic ancestry. Gen Res 13:624–634
13. Kittles RA, Bergen AW, Urbanek M et al (1999) Autosomal, mitochondrial, and Y chromosome DNA variation in Finland: evidence for a male-specific bottleneck. Am J Phys Anthropol 108:381–399
14. Lappalainen T, Koivumäki S, Salmela E et al (2006) Regional differences among the Finns: a Y-chromosomal perspective. Gene 376:207–215
15. Mayntz-Press KA, Ballantyne J (2007) Performance characteristics of commercial Y-STR multiplex systems. J Forensic Sci 52:1025–1034
16. Norio R (2003) Finnish Disease Heritage II: population prehistory and genetic roots of Finns. Hum Genet 112:457–469
17. Palo JU, Hedman M, Ulmanen I, Lukka M, Sajantila A (2007) High degree of Y-chromosomal divergence within Finland—forensic aspects. Forensic Sci Int Genet 1:120–124
18. Pereira RW, Monteiro EHG, Hirschfeld GCR, Wang AY, Grattapaglia D (2007) Haplotype diversity of 17 Y-chromosome STRs in Brazilians. Forensic Sci Int 171:226–236
19. Petit RJ, El Mousadik A, Pons O (1998) Identifying populations for conservation on the basis of genetic markers. Cons Biol 12:844–855
20. Redd AJ, Chamberlain VF, Kearney VF et al (2006) Genetic structure among 38 populations from the United States based on 11 U.S. core Y chromosome STRs. J Forensic Sci 51:580–585
21. Roewer L, Willuweit S, Krüger C et al (2008) Analysis of Y chromosome STR haplotypes in the European part of Russia reveals high diversities but non-significant genetic distances between populations. Int J Legal Med 122:219–223
22. Roewer L, Kayser M, de Knijff P et al (2000) 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 114:31–43
23. Roewer L, Croucher PJP, Willuweit S et al (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. Hum Genet 116:279–291

24. Sajantila A, Salem AH, Savolainen P, Bauer K, Gierig C, Pääbo S (1996) Paternal and maternal DNA lineages reveal a bottleneck in the founding of the Finnish population. P Natl Acad Sci USA 93:12035–12039
25. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
26. Underhill PA, Shen PD, Lin AA et al (2000) Y chromosome sequence variation and the history of human populations. Nat Genet 26:358–361
27. Underhill PA, Passarino G, Lin AA et al (2001) The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 65:43–62
28. Willuweit S, Roewer L (2007) Y chromosome haplotype reference database: Update. Forensic Sci Int Genetics 1:83–87