Hum Genet (2005) 116: 279–291
DOI 10.1007/s00439-004-1201-z

ORIGINAL INVESTIGATION

**Lutz Roewer · Peter J. P. Croucher · Sascha Willuweit
Tim T. Lu · Manfred Kayser · Rüdiger Lessig
Peter de Knijff · Mark A. Jobling · Chris Tyler-Smith
Michael Krawczak**

# Signature of recent historical events in the European Y-chromosomal STR haplotype distribution

Received: 13 May 2004 / Accepted: 13 September 2004 / Published online: 20 January 2005


**Abstract** Previous studies of human Y-chromosomal single-nucleotide polymorphisms (Y-SNPs) established a link between the extant Y-SNP haplogroup distribution and the prehistoric demography of Europe. By contrast, our analysis of seven rapidly evolving Y-chromosomal short tandem repeat loci (Y-STRs) in over 12,700 samples from 91 different locations in Europe reveals a signature of more recent historic events, not previously detected by other genetic markers. Cluster analysis based upon molecular variance yields two clearly identifiable sub-clusters of Western and Eastern European Y-STR haplotypes, and a diverse transition zone in central Europe, where haplotype spectra change more rapidly with longitude than with latitude. This and other observed patterns of Y-STR similarity may plausibly be related to particular historical incidents, including, for example, the expansion of the Franconian and Ottoman Empires. We conclude that Y-STRs may be capable of resolving male genealogies to an unparalleled degree and could therefore provide a useful means to study local population structure and recent demographic history.

L. Roewer and P.J.P. Croucher contributed equally to this paper.

L. Roewer · S. Willuweit
Institute of Legal Medicine, Humboldt-University, Berlin, Germany

P. J. P. Croucher
First Department of Medicine, Christian-Albrechts-University, Kiel, Germany

P. J. P. Croucher · T. T. Lu · M. Krawczak (✉)
Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Brunswiker Strasse 10, Kiel, Germany
E-mail: krawczak@medinfo.uni-kiel.de
Tel.: +49-431-5973200
Fax: +49-431-5973193

M. Kayser
Max-Planck-Institute for Evolutionary Anthropology, Leipzig, Germany

M. Kayser
Department of Forensic Molecular Biology, Erasmus University, Rotterdam, The Netherlands

R. Lessig
Institute of Legal Medicine, University of Leipzig, Germany

P. de Knijff
Forensic Laboratory for DNA Research, Leiden University, The Netherlands

M. A. Jobling
Department of Genetics, University of Leicester, UK

C. Tyler-Smith
The Wellcome Trust Sanger Institute, Hinxton, UK

## Introduction

The population dynamic processes that created the subtle gradation pattern of European culture, and eventually shaped the continent's genetic structure, are unparalleled in world history. It is widely accepted that extant Europeans have their earliest roots in the scattered Palaeolithic hunter-gatherer communities of around 40,000–43,000 years ago (Boyd and Silk 1997). Some 10,000 years before present (YBP), the rise of agriculture in the Near East led to the migration into Europe of a rapidly expanding farming population, although the extent of concomitant genetic change has been a matter of some dispute (Ammerman and Cavalli-Sforza 1984; Chikhi et al. 2002). Archaeological evidence suggests that the oldest rural villages in Europe developed on the western coast of the Aegean Sea and in

Crete around 7,000 YBP. From there, agriculture spread to most of the Balkan Peninsula and unfolded westwards via the Vardar-Danube-Rhine corridor and along the northern coast of the Mediterranean.

For thousands of years the South-East of Europe was the most developed part of the continent, whilst north of the Alps non-imperial and rather autonomous communities predominated. With the decline of the Roman Empire, around 2,000 YBP, the Teutons started to expand westwards, and the amalgamation of the Roman and Germanic civilisations marked the origin of the extant "occidental" European populations (Banniard 1989). By contrast, the eastern part of Europe always provided a gateway for migration and invasion by nomadic populations, from the Kurgan expansion around 7,000 YBP to the east-west movements of the Scythians, Mongols, Huns, Avars, Alanes and Magyars in historical times. Originating from the upper Dnjepr, the Slavs had settled on the southern shores of the Baltic Sea and replaced the Germanic speakers in Hungary and the Balkans during the sixth century AD (Gimbutas 1971). Their expansion was only halted by the new European powerhouse, the Franconian Empire. The political divide between these two cultures was marked by the rivers Elbe, Danube and Save, a split that has influenced Europe ever since and is still obvious today.

The population genetic consequences of the European colonisation and re-colonisation during Palaeolithic and Neolithic times have been addressed by a number of studies that utilised slowly evolving single-nucleotide polymorphisms (SNPs) on the Y chromosome (Rosser et al. 2000; Semino et al. 2000). Owing to its large number of polymorphic sites and its particular sensitivity to genetic drift (Seielstad et al. 1998; Kayser et al. 2001), the non-recombining part of the Y chromosome is especially useful for the investigation of population movements. Biallelic polymorphisms have proven useful for the analysis of prehistoric events, since the ancestral and derived sequence variants have had enough time to evolve independently. However, in order to investigate the impact of political, religious and cultural incidents in historical times, such as the so-called "Making of Europe" between 950 and 1350 AC (Bartlett 1994), greater resolution is required to distinguish the relevant patrilineal genealogies. The ability of hypervariable Y-chromosomal short tandem repeat (Y-STR) haplotypes to discriminate between even closely related or co-localised male populations has been demonstrated before for Germans and Dutch (Roewer et al. 1996), for the Baltic populations (Lessig et al. 2001), for Central England and North Wales (Weale et al. 2002) and for Poland and Germany (Ploski et al. 2002). In the present study, we genotyped over 12,700 Europeans at seven Y-STR loci and assessed whether the geographical haplotype distribution of these fast-evolving, male-specific markers indeed reflects recent history. Samples were derived from 91 different populations, spread across the culturally most diverse regions of the continent and representing the most extensive survey of human Y-chromosomal diversity undertaken in any population or group of populations to date.

## Material and methods

### DNA samples

DNA samples were obtained from 12727 "white" European males through 91 local recruitment units (Fig. 1). Nearly all individuals were ascertained via a judicial or private paternity case, and care was taken within the confines of general forensic practice that no closely related males or males of non-local origin were included in the study. The samples therefore represent an unbiased cross-section of that part of the respective male population that would have been liable to a paternity dispute. All samples were logged in the Y-STR haplotype reference database (YHRD), maintained at the Institute of Legal Medicine, Humboldt-University of Berlin, Germany, and made publicly available via the Internet. The names of the samples and meta-samples (see legend to Fig. 1) were chosen pragmatically without any prior causal or explanatory (e.g. linguistic) relationship in mind.

### Y-STR and Y-SNP genotyping

All samples were genotyped for six tetranucleotide Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS393) and one trinucleotide Y-STR (DYS392), following published protocols (Kayser et al. 1997). In accordance with recommendations by the International Society of Forensic Genetics (Gill et al. 2001), Y-STR alleles were designated according to the number of variable repeats included. Consistent allele designation was assured by the use of allelic ladders in all participating laboratories. Note that, in contrast to the common practice in forensic databases, our allele designation at DYS389I and DYS389II refers to the repeat number at individual loci, and not the repeat numbers revealed by the multiplex genotyping method employed (Rolf et al. 1998). Prior to inclusion in the study, all laboratories had to pass a quality test that involved blind genotyping of five control samples. Additional information about markers, laboratories and sampling procedures is available at the YRHD web site and from the Forensic Laboratory for DNA Research in Leiden, The Netherlands.

Genotyping of Y-chromosomal SNPs in the Baltic and Dutch samples allowed the assignment of each individual chromosome to one of the most common European Y-SNP haplogroups. Multiplex PCR and SNaPshot minisequencing reactions were carried out as described (Lessig et al. 2004). All products were analyzed on ABI Prism 310 or 3100 Avant Genetic Analyzers.

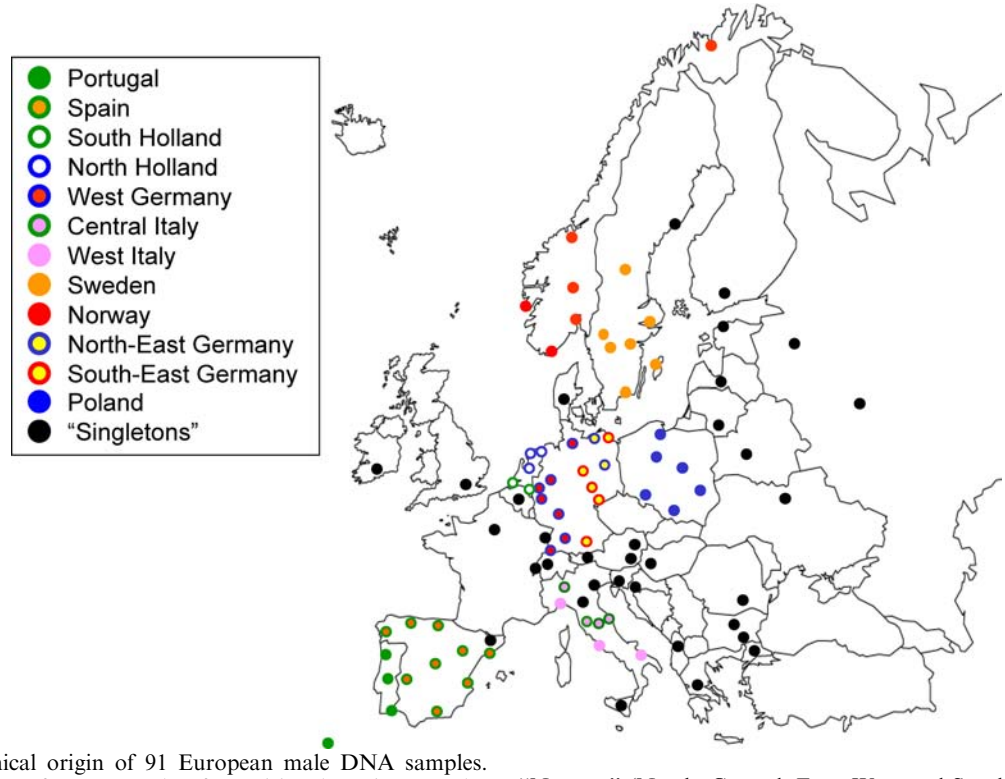


**Fig. 1** Geographical origin of 91 European male DNA samples. The colour coding of meta-samples, formed by clustering samples within a country according to minimum $\Phi_{ST}$, is given in the inset. *Black circles* mark singleton samples which were not included in any meta-sample. Meta-samples include "Spain" (Andalucia, Aragon, Asturias, Caceres, Cantabria, Catalonia, Galicia, Madrid, Valencia), "Portugal" (North, Central and South Portugal, Madeira), "South Holland" (Limburg, Zeeland), "North Holland" (Friesland, Groningen, Leiden), "Sweden" (Blekinge, Gotland, Ostergoetland, Skaraborg, Stockholm, Uppsala, Vaermland), "Norway" (North, Central, East, West and South Norway, Oslo), "West Germany" (Cologne, Düsseldorf, Freiburg, Hamburg, Mainz, Münster, Stuttgart), "North-East Germany" (Berlin, Rostock), "South-East Germany" (Chemnitz, Greifswald, Leipzig, Magdeburg, Munich), "Central Italy" (Lombardy, Marche, Tuscany, Umbria), "West Italy" (Latium, Liguria, Puglia), and "Poland" (Bydgoszcz, Gdansk, Krakow, Lublin, Warsaw, Wroclaw)

## Spatial autocorrelation analysis

Spatial autocorrelation analysis of haplotypes was performed using Moran's *I* index (Sokal and Oden 1978). Parameter *I* measures the correlation between observations of the same type, made at locations of a given geographical distance. Here, we used the repeat number of each Y-STR locus as the primary observation. The pair-wise great circle distances (GCD) between haplotype origins (i.e. recruitment units) were determined using Cartesian coordinates obtained from the 2004 version of the CIA World Factbook and other sources in the public domain. We chose not to employ *II*, an allele frequency-based adaptation of Moran's *I* to genetic analyses suggested by Barbujani (1987), since STR repeat numbers evolving through single-step mutation represent suitable quantities for analysis on their own. Autocorrelation was evaluated over 11 distance classes, including one class for haplotypes found in the same location (i.e. GCD equal to zero) and ten additional, equally frequent distance classes. Moran's *I* values were assessed for statistical significance using a randomisation test that permuted samples over locations in 100 replications.

## Analysis of molecular variance

The genetic relationship between different populations was assessed by means of $\Phi_{ST}$, an analogue of Wright's $F_{ST}$ that takes the evolutionary distance between individual haplotypes into account (Excoffier et al. 1992; Excoffier and Smouse 1994). Estimates of $\Phi_{ST}$ were obtained using the Arlequin software (Schneider et al. 2000) and tested for statistical significance by means of randomisation (1,000 replicates per comparison). Several local population comparisons yielded exceptionally low $\Phi_{ST}$ values, suggesting that the respective samples were genetically indistinguishable. Population samples were, therefore, recursively clustered into meta-samples, based upon $\Phi_{ST}$. In each clustering step, that pair of populations or clusters that yielded the minimum $\Phi_{ST}$ value was grouped together. Clustering was performed in a two-tiered fashion, in that groupings were first confined to within one and the same country, and grouping continued until no further insignificant clustering (i.e. $P > 0.05$ for the respective $\Phi_{ST}$) was possible within any country. This procedure led to the definition of regional meta-samples (Fig. 1) which were more comparable to samples from countries with only one recruitment unit. Meta-samples

and the remaining individual samples formed the basis of a second, exhaustive round of $\Phi_{ST}$-based clustering.

## Multidimensional scaling analysis

In order to graphically visualise the Y-STR genetic landscape of Europe, principle coordinates were identified by subjecting the pair-wise $\Phi_{ST}$ estimates between all 91 samples to a multidimensional scaling analysis (MDS) using SPSS (SPSS, Chicago Ill., USA). Individual solutions for one through six dimensions were iterated until the improvement in stress was less than 0.0001. The optimum dimensionality was then determined from a "scree" test (Table 1). A clear "elbow" was observed for the three-dimensional solution, with higher-dimensional solutions not providing a substantial decrease in Kruskal's stress value, $S$. The $R^2$ values (Table 1) indicate that 96.4% of the variance in the data can be accounted for by the three-dimensional solution.

**Table 1** Multidimensional scaling and "scree" test of pair-wise $\Phi_{ST}$ values ($S$ Kruskal's stress value, $R^2$ coefficient of determination)

| No. of dimensions | $S$ | $R^2$ |
|---|---|---|
| 1 | 0.211 | 0.887 |
| 2 | 0.136 | 0.944 |
| 3 | 0.106 | 0.964 |
| 4 | 0.094 | 0.968 |
| 5 | 0.084 | 0.972 |
| 6 | 0.076 | 0.975 |

The resulting coordinates were then used to generate three independent maps, one for each dimension, in which MDS coordinates were interpolated across Cartesian coordinates by inverse distance squared weighting with the 12 nearest neighbour samples, using the GRASS (Geographical Resources Analysis Support System) software. The European land mass was masked from oceans and seas using the GTOPO30 (30-arc seconds topographical) dataset, publicly available at the US Land Processes Distributed Active Archive Center. For better resolution, some sample locations were moved to the centre of the respective country of origin, especially when only one location was included for that country (e.g. Great Britain, Finland, Hungary etc.).

The first dimension of the MDS revealed a strong east-west gradient for the Y-STR genetic structure. In order to corroborate this finding more systematically, Spearman rank correlation coefficients were calculated for each of the 91 original samples between pair-wise $\Phi_{ST}$ and the longitudinal and latitudinal geographical distance, respectively, to all other samples. Singletons and meta-samples were also subjected to a pseudo-admixture analysis relative to the "Western Europe" and "Eastern Europe" sub-clusters that become clearly identifiable in the $\Phi_{ST}$-based cluster analysis (*boxed* in Fig. 2). Each haplotype that occurred in at least one of the two fringe sub-clusters was labelled either "Western" or "Eastern", depending upon where it was relatively more frequent. All haplotypes that were one-step neighbours of an Eastern or a Western haplotype of at least 1% frequency in the respective sub-cluster were assigned the same origin. All other European samples

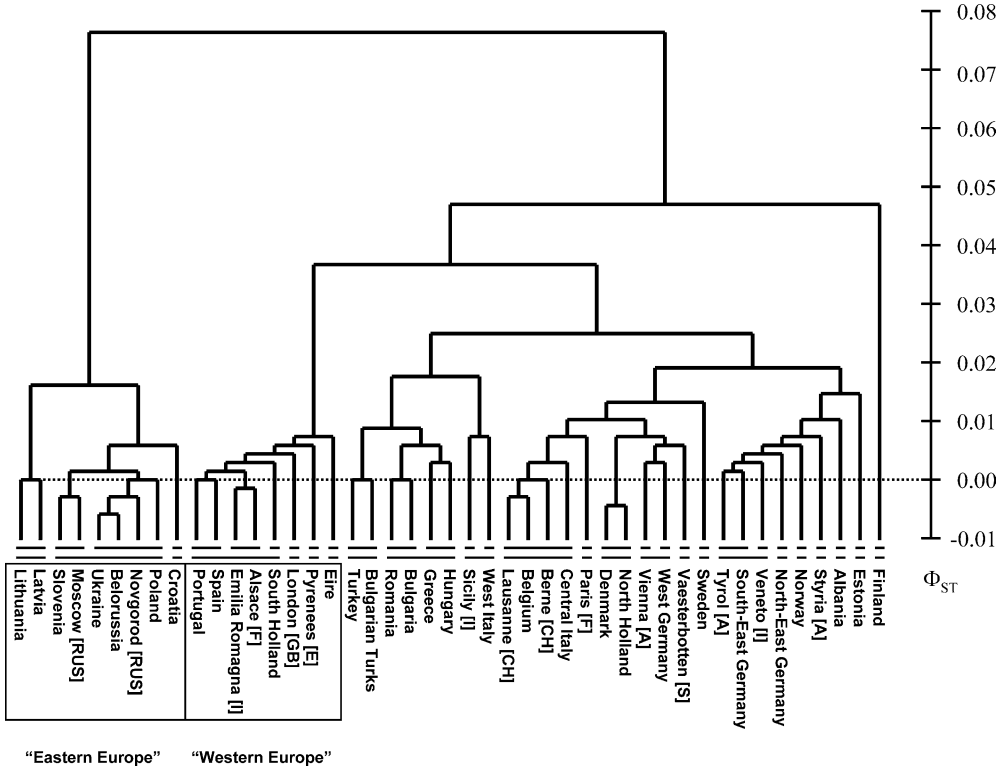


**Fig. 2** Clustering by minimum Y-STR-based $\Phi_{ST}$ of 45 male European samples and meta-samples. The significance of individual groupings is indicated by *vertical bars* below the dendrogram (*top line* randomisation $P < 0.05$, *bottom line* $P < 0.001$)

and meta-samples were then characterised in terms of the relative proportion of the fringe haplotypes.

### Assessment of local haplotype diversity variation

Under an infinite sites model (ISM) of mutation, the number of different haplotypes, $S_n$, expected in a population sample of size $n$ is related to the mutation rate $\mu$ and the effective population size $N$ via Ewen's sampling formula (Ewens 1972)

$$\mathrm{E}(S_n) = \sum_{i=0}^{n-1} \frac{\theta}{\theta + i}, \quad (1)$$

where $\theta = N\mu$ for haploid genomes. When the mutational process generating Y-STR haplotypes is approximated by an ISM (Helgason et al. 2000), adoption of $\mu = 2\times10^{-2}$ for all seven loci combined (Kayser et al. 2000) allows the estimation of effective population sizes from $n$ and $S_n$, solving Ewen's sampling formula for $\theta$. Furthermore, for any two populations with constant population sizes $N_1$ and $N_2$, respectively, and a constant number $M$ of migrants per generation between them, assuming that each migrant introduces a new haplotype into the other population leads to expected values, $E(S_{n,\,k})$, that correspond to

$$\theta_k = N_k\left(\mu + \frac{M}{N_k}\right), \quad k = 1, 2. \quad (2)$$

For the combined population, $E(S_{n,\mathrm{c}})$ is such that

$$\theta_\mathrm{c} = (N_1 + N_2)\mu \quad (3)$$

if the two populations are completely panmictic. If not, then $\theta_\mathrm{c}$ as obtained from $S_{n,\mathrm{c}}$ using formula 1 an unbiased estimate of the right-hand side of formula 3 only if $n_1/n_2 = N_1/N_2$, i.e. if the two sample sizes are proportional to the population sizes. However, even under complete population separation (i.e. $M = 0$), simulation has shown that the respective bias is not very serious for a wide range of $n_i$ values since the curve that relates $\theta_\mathrm{c}$ to $n_1$ and $n_2$ is usually flat around $n_1/n_2 = N_1/N_2$ (data not shown).

## Results

### Populations, haplotypes and genetic diversity

The Y-chromosomal STRs DYS19, DYS389I, DYS389-II, DYS390, DYS391, DYS392 and DYS393 were analysed in 12,727 males from 91 European populations (Fig. 1). In the total sample, the markers showed between eight (DYS391, DYS393) and 11 (DYS390, DYS392) different alleles. However, of the seven million different theoretically possible haplotypes, only 2,489 were actually observed. More than half of these (1,397/2,489 = 56.1%) were unique to an individual male. Fifteen haplotypes, comprising 3,088 males (24.3%), occurred in 40 or more populations; another 160 haplotypes, representing 4,692 individuals (36.9%), were found in 10–39 populations. The most frequent haplotype (''14-13-16-24-11-13-13'', in the above marker order) accounted for 661 males (5.2%) and was observed in 80 populations. Its 14 one-step neighbours comprised another 1,368 individuals (10.7%), with observed haplotype counts ranging from 24 to 303. Whilst the highest haplotype diversity $h$ (Nei 1987) was observed in Vienna, where all 66 haplotypes were different ($h = 1.00$), particularly low values of $h$ (i.e. $h < 0.95$) were noted in Albania and Finland.

When population samples were iteratively collapsed within countries, a total of 12 meta-samples emerged before all $\Phi_\mathrm{ST}$ values became significant at the 5% level (Fig. 1). This regrouping involved 58 original samples; the remaining 33 data sets were henceforth treated as singletons. The second, exhaustive round of clustering by minimum $\Phi_\mathrm{ST}$ clearly reflected the geographical actuality of Europe (Fig. 2). All Western populations, with the exception of metropolitan Paris, formed a clearly identifiable sub-cluster (''Western Europe''). This sub-cluster also included Emilia Romagna from Italy, which in turn strongly resembled the French sample from the Alsace ($\Phi_\mathrm{ST} = -0.0010$, $P = 0.507$). Most notably, the Spanish and Portuguese samples were not found to be significantly different ($\Phi_\mathrm{ST} = 0.0014$, $P = 0.078$). At the other extreme, all Slavs except Bulgaria formed one cluster (''Eastern Europe''), where only Croatia ($\Phi_\mathrm{ST} = 0.0061$) and the two Baltic samples ($\Phi_\mathrm{ST} = 0.0157$) differed from the rest at the 0.1% significance level. Bulgaria belonged to a different sub-cluster, which also incorporated Romania ($\Phi_\mathrm{ST} = 0.0001$, $P = 0.417$), and which extended to Greece and Hungary at the 1% significance level ($\Phi_\mathrm{ST} = 0.0055$, $P = 0.013$ for all four samples combined). Sicily and ''West-Italy'' together with the two Turkish samples clearly complemented this ''Balkan-Danube'' grouping to form a coherent ''Southern European'' sub-cluster. Whilst the two southern Dutch populations (''South Holland'') were included in ''Western Europe'', the three northern samples (''North Holland'') formed a ''Friesian'' sub-cluster together with Denmark ($\Phi_\mathrm{ST} = -0.0044$, $P = 0.795$). The Friesians were part of a more comprehensive sub-cluster, including all francophone populations, West Germany, Vienna and Sweden, which may be loosely called ''West-Central Europe''. By contrast, the other German and Austrian samples as well as Northern-Italian Veneto link in with Norway and eventually Albania and Estonia to group as a ''East-Central Europe''. Finally, Finland can be regarded as a sub-cluster on its own which associated with the non-Slavic populations ($\Phi_\mathrm{ST} = 0.0460$, $P < 0.001$) before both connected with the ''Eastern European'' sub-cluster ($\Phi_\mathrm{ST} = 0.0744$, $P < 0.001$).

### Genetic and geographic distances

Spatial autocorrelation analysis yielded results that were consistent with an ''isolation by distance'' model of

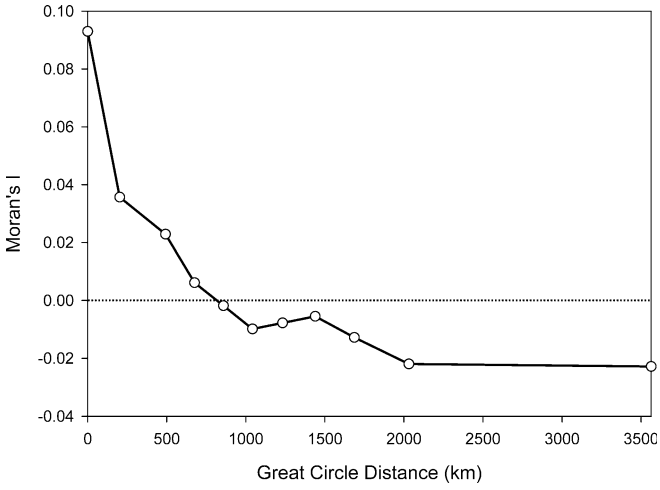


**Fig. 3** Spatial autocorrelation analysis of European Y-STR haplotypes. Moran's *I* index was calculated from the repeat number at each Y-STR locus, using the Great Circle Distance between the respective recruitment units as a measure of pair-wise geographical distance between haplotypes

haplotype divergence (Fig. 3). Thus, Y-STR repeat numbers were most similar for haplotypes sampled in the same location ($I = 0.0929$). At less than 1,000 km, the autocorrelation index decreased sharply with distance, reaching a plateau of negative *I* values thereafter and decreasing only slightly beyond 2,000 km. All indices were significantly different from zero (randomisation $P < 0.01$), including $I = -0.0018$ for the distance class centred at 940 km.

Inspection of Fig. 1 suggests that the similarity of Y-STR haplotypes decays much more rapidly along an east-west than a north-south gradient, at least in central Europe. This notion was formally corroborated by a MDS analysis of all pair-wise $\Phi_{ST}$ values. The first dimension, accounting for almost 89% of the variance (Table 1), clearly shows a decomposition of the European Y-STR genetic structure into three major components (Fig. 4a), closely corresponding to the "Western", "Central" and "Eastern" sub-clusters of Fig. 2. The first dimension also highlights the genetic peculiarity of metropolitan Paris and Vienna, Finland, and the two Balkan-Slavic samples of Slovenia and Croatia in relation to their respective surroundings. The degree of east-west stratification of the European Y-STR haplotype spectrum was quantified by Spearman rank correlation analysis between the latitudinal and longitudinal distances, respectively, and pair-wise $\Phi_{ST}$ (Fig. 5). For 81 samples, the correlation was stronger with longitude than with latitude, and the few populations showing a notably reversed effect were from the fringe of the

►

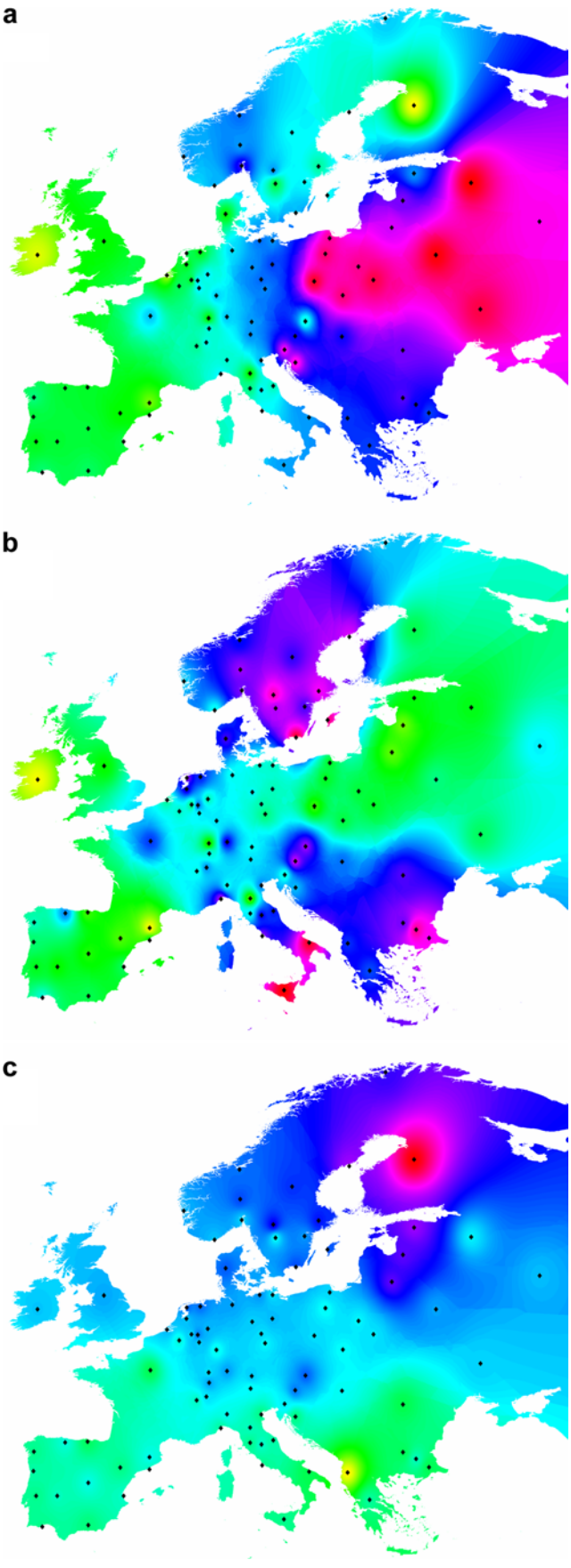


**Fig. 4a–c** Multidimensional scaling analysis of pair-wise Y-STR-based $\Phi_{ST}$ between 91 male European samples. Displayed are the first three dimensions (**a**–**c**) which together account for 96.4% of the variance. Sample locations are marked in *black*; colour coding is on an arbitrary "rainbow" scale that allocates *yellow* and *magenta* to the opposite extremes, via *green* and *blue*

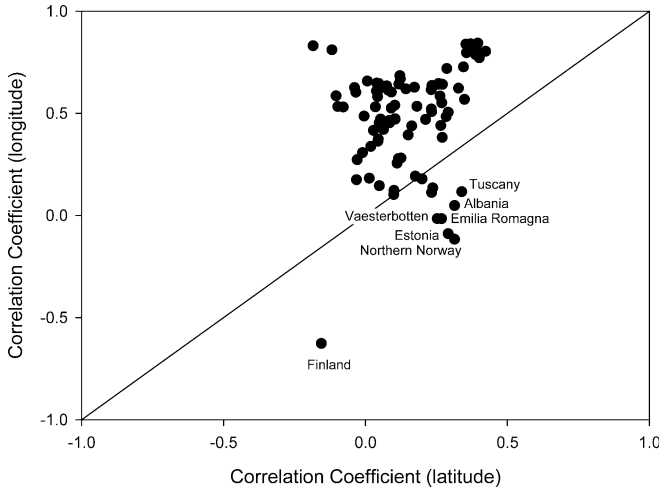


**Fig. 5** Longitudinal and latitudinal extent of relatedness among European Y-STR haplotypes. *Each dot* represents one of 91 recruitment units and depicts the Spearman correlation coefficient of the pair-wise $\Phi_{ST}$ values between that unit and all other samples with the respective geographical distances. *Horizontal axis* correlation with latitudinal distances (north–south), *vertical axis* correlation with longitudinal distances (east–west)

continent (Fig. 5). Furthermore, whilst only five samples showed a negative correlation with longitude, namely Emilia Romagna (I), Vaesterbotten (S), Finland, Estonia and Northern Norway, the same was true for 12 samples with latitude. The second dimension of the MDS analysis revealed more subtle structural features, such as, for example, the distinction between the Turkish and non-Turkish samples in "Southern Europe" and the divide between the two Dutch meta-samples. The third dimension eventually depicted an underlying north-south gradient that is usually seen in Y-SNP studies of European populations (Rosser et al. 2000; Semino et al. 2000). However, since the second and third dimensions of the Y-STR MDS accounted for less than 10% of the variance (Table 1), the major geographic structuring associated with the two types of markers must be substantially different.

## Pseudo-admixture analysis

When compared with the peripheral "Western Europe" and "Eastern Europe" sub-clusters, the other 28 samples/meta-samples exhibited prominent geographical gradients in terms of their Y-STR haplotype spectra (Table 2). The proportion of Western European haplotypes was naturally highest in "Western Europe" itself (91%), with a cline-like decrease as one moves east and north from Belgium (66%) and North Holland (61%) to Romania (25.0%), Estonia (14%),"Eastern Europe"

**Table 2** Pseudo-admixture analysis of Y-STR haplotypes in European populations (*n* total number of haplotypes, *NA* not applicable)

| Sample/meta-sample | *n* | Frequency of haplotype group (%) | | |
|---|---|---|---|---|
| | | Western | Eastern | None |
| Western Europe | 2,529 | 2300 (91) | 229 (9) | NA |
| Finland | 399 | 39 (10) | 266 (67) | 94 (23) |
| Estonia | 133 | 19(14) | 85 (64) | 29(22) |
| Albania | 101 | 34 (34) | 54 (53) | 13 (13) |
| Styria (A) | 65 | 19 (29) | 16 (25) | 30 (46) |
| Norway | 300 | 93 (31) | 151 (50) | 56 (19) |
| North-East Germany | 752 | 268 (35) | 321 (43) | 163 (22) |
| Veneto (I) | 120 | 62 (52) | 18 (15) | 40 (33) |
| South-East Germany | 1,404 | 613 (44) | 551 (39) | 240 (17) |
| Tyrol (A) | 229 | 100 (44) | 88 (38) | 41 (18) |
| Sweden | 667 | 233 (35) | 304 (46) | 130 (19) |
| Vaesterbotten (S) | 41 | 13 (32) | 20 (48) | 8 (20) |
| West Germany | 1,287 | 631 (49) | 418 (32) | 238 (19) |
| Vienna (A) | 66 | 19 (29) | 14 (21) | 33 (50) |
| North Holland | 179 | 110 (61) | 43 (24) | 26 (15) |
| Denmark | 63 | 35 (56) | 17 (27) | 11 (17) |
| Paris (F) | 109 | 51 (46) | 17 (16) | 41 (38) |
| Central Italy | 559 | 316 (57) | 80 (14) | 163 (29) |
| Berne (CH) | 91 | 47 (52) | 24 (26) | 20 (22) |
| Belgium | 125 | 83 (66) | 22 (18) | 20 (16) |
| Lausanne (CH) | 108 | 64 (60) | 22 (20) | 22 (20) |
| West Italy | 373 | 193 (52) | 75 (20) | 105 (28) |
| Sicily (I) | 199 | 70 (35) | 36 (18) | 93 (47) |
| Hungary | 118 | 35 (30) | 62 (52) | 21 (18) |
| Greece | 101 | 45 (44) | 27 (27) | 29 (29) |
| Bulgaria | 122 | 34 (28) | 65 (53) | 23 (19) |
| Romania | 102 | 25 (24) | 58 (57) | 19 (19) |
| Bulgarian Turks | 61 | 15 (25) | 30 (49) | 16 (26) |
| Turkey | 158 | 54 (34) | 46 (29) | 58 (37) |
| Eastern Europe | 2,166 | 228 (11) | 1938 (89) | NA |
| Total | 12,727 | 5,848 (46) | 5,097 (40) | 1,782 (14) |

(11%) and Finland (10%). The proportion of Eastern European haplotypes showed a similar trend in the reverse direction. Some samples were characterised by a relatively high proportion of haplotypes (>30%) that were not classifiable as either "Eastern" or "Western", including Vienna (50%), Sicily (47%), Styria (46%), Paris (38%), Turkey (37%) and Veneto (33%).

### Local variation in the Dutch and Baltic Y-STR haplotype spectra

Since the effective population size estimates for the North and South Holland meta-samples were neither additive nor identical (Table 3), the two sub-populations cannot have been completely separated or panmictic. Instead, they must have experienced some level of recent migration, and solving formulae 2 and 3 with the respective $\theta$ values included yielded $N_1 = 3{,}918$, $N_2 = 1{,}298$ and $M = 22$. These estimates were verified by 10,000 simulations of an ISM with migration, using the very same $N_1$, $N_2$ and $M$ values. The mean $S_n$ obtained in simulated samples of the original sizes (i.e. $n_1 = 179$, $n_2 = 96$, and $n_3 = n_1 + n_2 = 275$) were $100.3 \pm 6.1$, $52.3 \pm 4.2$ and $136.5 \pm 7.8$, and were therefore very close to the actually values (Table 3). As a consequence, it may be concluded that the Dutch Y-STR data are best explained by a threefold higher effective population size in North than in South Holland and by migration rates of $m_1 = 22/3918 = 0.56\%$ (in and out of the North), and $m_2 = 22/1298 = 1.69\%$ (in and out of the South). On the other hand, over the last, say, 100 generations, the Dutch population can be assumed to have been large enough for genetic drift to have had a minor effect upon its genetic structure. Under this assumption, the change in frequency of a given allele or haplotype can be modelled in the two sub-populations as

$$|f_1(t) - f_2(t)| = (1 - m_1 - m_2)^t \cdot |f_1(0) - f_2(0)| \qquad (4)$$

For the Dutch population, this implies that any difference in allele or haplotype frequency would have decreased by $0.56 + 1.69 = 2.25\%$ in one generation, and by $100 \cdot [1 - (1 - 0.0225)^{100}] = 89.7\%$ in 100 generations.

**Table 3** Patterns of Y-STR haplotype diversity in the Dutch and Baltic populations ($n$ total number of haplotypes, $S_n$ number of different haplotypes, $\theta$ population parameter, $N$ effective population size)

| Sample | $n$ | $S_n$ | $\theta$ | $N$ |
|---|---|---|---|---|
| North Holland | 179 | 103 | 100.2 | 5,010 |
| South Holland | 96 | 53 | 47.8 | 2,390 |
| Combined | 275 | 135 | 104.3 | 5,215 |
| Estonia | 133 | 93 | 136.0 | 6,800 |
| Latvia/Lithuania | 296 | 171 | 168.0 | 8,400 |
| Combined | 429 | 229 | 198.9 | 9,945 |

Under the same migration/ISM, we obtained $N_1 = 4{,}172$, $N_2 = 5{,}772$, and $M = 53$ for the two Baltic samples/meta-samples from Estonia ("1") and Lithuania/Latvia ("2"), respectively (Table 3). This implies that the migration rates between the two Baltic sub-regions must have been of a similar order ($m_1 = 0.0127$, $m_2 = 0.0092$) as those between the Dutch sub-populations.

## Discussion

The genealogy of European Y chromosomes is characterised by substantial differentiation (Rosser et al. 2000; Semino et al. 2000) and, as expounded upon in more detail below, our study suggests that the observed pattern of cline-like similarity between Y-chromosomal STR haplotypes may partly reflect demographic events that occurred in historical time. Recent male demography would indeed be best assessed by the analysis of Y-chromosomal polymorphisms since the fourfold smaller effective population size of male-to-male transmission renders Y-chromosomal markers much more sensitive to genetic drift and population bottlenecks than their autosomal counterparts. In principle, similar arguments should also apply to mitochondrial (mtDNA) markers, but little geographical structuring has been detected in Europe by means of mtDNA analysis (Richards et al. 2002). This has led to the suggestion that mtDNA is either exceptionally selection-sensitive or that female gene flow has been particularly high in Europe (Barbujani and Chikhi 2000).

Y-chromosomal short tandem repeats, in particular, are capable of resolving population strata into individual genealogies that would otherwise be inseparable (Roewer et al. 1996; Lessig et al. 2001; Ploski et al. 2002; Weale et al. 2001, 2002). The power of the Y-STR approach is, however, critically dependent upon careful sample collection, especially in regions with neighbouring populations that have undergone mutual transitions. Our analysis has made opportunistic use of the Y-STR haplotype reference database (YHRD), a survey of male genetic variation in Europe that was initiated in 1997 by the Forensic Y Chromosome Research Group. The major purpose of the database was to serve as reference material for the presentation of male DNA profiles in court (Roewer et al. 2000). Since then, the repository has been growing continuously and expanded beyond its original scope. As of August 2004, YHRD contained over 24,000 Y-STR haplotypes from over 230 populations world-wide. The history of the database and its originally intended use in forensic practice imply that sampling has been neither systematic nor comprehensive. However, it is questionable whether a merely academic study of neutral population genetic variation on the present scale would have ever been possible to implement. Furthermore, the use of validated genotyping procedures in quality-controlled forensic laboratories has served to ensure the highest credibility for each

and every haplotype that entered into YHRD (Roewer et al. 2001).

In contrast to previous, SNP-based, studies of European male genetics (Rosser et al. 2000; Semino et al. 2000), our analysis has utilised a much higher level of inter-individual variability, and the identification of previously unrecognised male strata in Europe confirms the utility of this approach. However, the high variability of Y-STR profiles is a direct consequence of the $10^5$ to $10^6$-fold higher likelihood of meiotic repeat length change than single base-pair substitution (Kayser et al. 2000). High mutation rates may homogenise haplotypes of different descent (de Knijff 2000), an inherent drawback of STR haplotypes that could only be compensated for in our study by extensive sampling. The large size of most regional samples and meta-samples ensured that all locally common patrilines were represented by haplotype classes that contain profiles separated by one or only a few single-step mutations.

Our study provides clear evidence for a major genetic division of European males into Slavic-speaking eastern and Romance language-speaking western populations, separated by a central European block of Germanic- and Italian-speaking populations. The Western and Eastern European geographical clusters are reminiscent of the well-established distribution of Y-SNP haplogroups P(xR1a) and R1a (Rosser et al. 2000). Indeed, preliminary data by Kittler et al. (2003) indicate that the ancestral Y-STR haplotypes were "14-13-16-24-11-13-13" for P(xR1a) and "16(?)-13-17-25-10-11-13" for R1a. Since the respective frequencies of these two haplotypes and their one-step neighbours were 30.6 and 0.6% in "Western Europe", as opposed to 4.4 and 20.5% in "Eastern Europe", the Y-STR-based and Y-SNP-based power of discrimination between the two sub-clusters would have been equivalent. Furthermore, the significantly negative autocorrelation in Y-STR repeat number observed for haplotypes located more than 1,000 km apart (Fig. 3) is suggestive of a continent-wide cline between at least two different ancient lineages, possibly corresponding to P(xR1a) and R1a. It is important to stress, however, that the autocorrelation was based on allele length, and not frequencies, so that we prefer to interpret Fig. 3 in broader terms as simply representing "isolation-by-distance".

Our data revealed a rapid change in haplotype spectra along east-west gradients, but a relative constancy over more than 1,000 km in a north-south direction in central parts of the continent. This may be indicative of the fact that, when the expanding Slavic and Franconian spheres of influence met in medieval central Europe, the only way for male lineages to expand further may have been in a northerly or southerly direction. Interestingly, in our pseudo-admixture analysis, even the male lineages within contemporary Germany turned out to be notably different in the "North-East" ($n = 752$; 43% Eastern vs 35% Western), the "South East" ($n = 752$; 39% Eastern vs 44% Western) and the "West" ($n = 1{,}287$; 32% Eastern vs 49% Western). When samples are stratified in this way, the difference between the three Y-STR haplotype spectra is highly significant ($\chi^2 = 3.945$, 2 *df*, $P < 0.001$). The area covered by the former German Democratic Republic significantly overlaps with the homeland of Slavic (i.e. Wendish) people from the Middle Ages, including the Sorbes, Pomeranes, Wagriens, Obodrites, and Ranes. This geographical coincidence would explain the obvious preservation of "Slavic" haplotypes in eastern Germany far better than, for example, the settlement of eastern European World War II refugees, since the latter were mostly Germans anyway.

In the Netherlands, a remarkable subdivision became apparent in that the two southern samples ($n = 96$) were included in the "Western Europe" sub-cluster whereas the three northern samples ($n = 179$) formed part of a "Friesian" sub-cluster, together with Denmark. The significant difference between the two Dutch regions was exclusively due to haplotypes "14-13-16-24-10-13-13" and "14-13-16-24-11-13-13". Their combined frequency was $16/179 = 0.089$ in the North and $23/96 = 0.240$ in the South ($\chi^2 = 1.583$, 1 *df*, $P < 0.001$). If the two haplotypes, which are also the most frequent ones in Spanish males, are excluded from both meta-samples, pair-wise $\Phi_{ST}$ drops from a highly significant 0.0183 ($P = 0.003$) to an insignificant 0.0069 ($P = 0.086$). In the light of the 1–2% background migration implied by the general overlap in Y-STR haplotype spectra, however, such an isolated and focused frequency difference must have had a fairly recent origin for it not to have been eradicated. One possible explanation for this coincidence would be a war-related instance from the 15th and 16th centuries when the south of the Netherlands came under Burgundian and Spanish control much earlier than the northern provinces. The latter formed the Union of Utrecht in 1579 and gained independence from Spain earlier than the southern Union of Arras (Merriman 1996).

Our considerations about the origin of Dutch Y-STR differences have been based upon various simplifying assumptions, the validity of which is not, however, critical for our main conclusion. First, we have adopted an ISM which is known to overestimate the number of STR alleles (or haplotypes) generated by step-wise mutation (Shriver et al. 1993). In part, this can be corrected for by adopting smaller $\theta$ values in the ISM; i.e. smaller population sizes and/or a smaller mutation rate. However, even if we assume $\mu = 1{\times}10^{-2}$ for the seven Y-STRs combined, this leads to $N_1 = 7{,}836$, $N_2 = 2{,}596$, $M = 22$, and $m_1 = 0.0028$, $m_2 = 0.0085$. Such a level of background mutation is still incompatible with the persistence over more than 100 generations of a 8.9 vs 24.0% allele frequency difference for a Y-SNP, and even more so for a Y-STR. Second, we have neglected migration from outside the Netherlands. Since this can be allowed for by increasing the mutation rate, the effective population sizes would have to be even smaller and the relative background migration even higher than before.

Similarly, we have assumed that every migrant introduces a new haplotype. Since this is clearly untrue, even under an ISM, the actual level of migration must have been even higher than estimated in order to achieve the observed effects upon $\theta$; i.e. an apparent increase in size over that expected if the two regions had been completely separated. Third, we assumed constant migration and a constant population size, whereas in reality the two populations would have undergone exponential expansion. This simplifying assumption would have biased the interpretation of the haplotype pattern in favour of our hypothesis only if the two regions had experienced a lot of migration occurring between them when populations were small, and became completely separated only more recently. If anything, European history suggests the opposite to be true. Finally, it can be shown that the difference between the two Dutch sub-regions is not detectable with Y-SNPs. SNP haplotype data were available for 171 (North) and 87 (South) of the males included in the Y-STR study, respectively, and the pair-wise $\Phi_{ST}$ value obtained with the discriminating Y-SNPs resolving the relevant European haplogroups was 0.0041 ($P = 0.209$).

In western Europe, Y-STRs identified a large coherent population pool, covering the Iberian Peninsula, parts of Italy, France (Alsace) and the British Isles, which coincides with the Frankonian Empire. Being the heiress of multiethnic Rome, ''Latin Europe'' (i.e. the part of Europe that was originally Roman Catholic rather than Greek orthodox or non-Christian) formed a zone where strong shared cultural features were as important as geographical contrasts (Bartlett 1994). Italian- and Germanic-speaking populations were integrated by adopting a common culture that was further consolidated by a strong political system. The eastern part of Europe, also identified by Y-STRs as being homogeneous in terms of its male genetic make-up, has a completely different history. This part of the continent was significantly influenced by various waves of immigration by nomadic Asian populations (Rosser 2000; Wells et al. 2001). Nevertheless, in the vast and hardly structured territory covered by our analysis, with no borders to the east, the Slavic language was clearly a strong uniting element for diverse cultures. The colonisation attempts of western civilisations appear to have left no significant genetic traces in eastern Europe and, during the 20th century, genocide and resettlement have both served to homogenise the (formerly more varied) genetic landscape even further (Ploski et al. 2002).

In the Baltics, our analysis confirmed that Estonian male lineages are comparatively close to both the Central Europeans and Finnish, but differ strikingly from the Lithuanian and Latvian Y-STR haplotype spectra (Lessig et al. 2001). The difference between the two samples/meta-samples is partly attributable to haplotypes ''14-14-16-23-11-14-14'' and ''14-14-16-24-11-14-14'' which occurred with frequencies of 14/129 = 0.109 in Estonia and 4/296 = 0.014 in the other two countries (Fisher's exact two-sided $P = 3.1 \cdot 10^{-5}$). These two haplotypes are also the most frequent ones among Finish males. With one exception (''R1a1'' on a Latvian chromosome), they were found to be associated with Y-SNP haplogroup ''N3'' in the Baltic samples. Haplogroup ''N3'' is indeed more frequent in Estonia (49/129 = 0.380) than in Latvia and Lithuania (80/296 = 0.270; $\chi^2 = 5.103$, 1 *df*, $P = 0.024$) but not when the above haplotypes are excluded (35/115 = 0.304 vs 77/292 = 0.264; $\chi^2 = 0.648$, 1 *df*, $P > 0.4$). Interestingly, the most frequent Estonian haplotype is ''14-12-16-22-10-11-13'', with a frequency of 9/129 = 0.070 versus 1/296 = 0.003 in Latvia and Lithuania (Fisher's exact two-sided $P = 1.3 \cdot 10^{-4}$). This haplotype is the third most frequent one in Germany, and all but one copy in the Baltic samples were found to reside on SNP haplogroup ''I''. Nevertheless, the Y-STR haplotype frequency difference is not explicable in terms of a generally higher frequency of haplogroup ''I'' in Estonians. Excluding the ten Germanic Y-STR haplotypes in question, the frequencies of haplogroup ''I'' are virtually identical (15/120 = 0.125 in Estonia and 33/295 = 0.112 in Latvia/Lithuania). The most likely explanation for the substantial division exhibited by the Baltic male lineages is the linguistic barrier that separates the Lithuanians and Latvians, who speak two of the three Baltic languages, and the Estonians, whose language is part of the Finno-Ugric group. The similarity of the Lithuanian and Latvian Y-chromosomal gene pool, in turn, is potentially explicable by the fact that a large proportion of present-day Latvia came under Lithuanian-Polish control in 1561 and remained so until the second division of Poland in 1772 (Gieysztor et al.1979).

The sub-cluster broadly described as ''Balkan-Danube'' comprises Greek, Romanian, Bulgarian and Hungarian males, i.e. members of populations who speak languages from very different families. The ''Balkan-Danube'' sub-cluster is also a relatively short distance away from the ''Turks'' (Fig. 2) and relatively distant to the eastern (Slavic) samples. This pattern of genetic similarity may seem counterintuitive at first glance since the geographical region of origin of the ''Balkan-Danube'' meta-sample overlaps considerably with the territory occupied by the southern Slavs after their separation from the eastern and western Slavic populations in the fifth and sixth centuries (Gimbutas 1971). In the 16th century, however, the same territory was ruled by the Turks during the Ottoman expansion and, 1,000 years earlier, Finno-Ugric-speaking (Magyars) and Turkic-speaking people (e.g. the Proto-Bulgarians) from western Asia had already invaded and settled in the Danube basin and the Balkans, and had consequently amalgamated with the southern Slavic populations present. Another important factor that divides the Slavic world has been religious orientation, and even today one of the sharpest cultural divisions in the Slavic world is that between people converted to Christianity by the Franks and the Greeks, respectively (Bartlett 1994). More than anywhere else, the genetic

record of the ''Balkan-Danube'' region can thus be read as a palimpsest of repeatedly ''overwritten'' historical processes (Jobling et al. 2004).

The appearance of singletons which are clearly genetically different from the surrounding populations is a result of sampling either from highly admixed and heterogeneous Metropolitan populations (Paris, Vienna), or sampling populations from the fringes of Europe that have a very distinct history (Finns, Estonians, Irish, Albanians). For the latter category, the most important force creating genetic individuality would have been patrilocality, a socially instituted practice whereby a newly-married couple live with or near the family of the husband. The importance of social (family) structure in influencing the population genetic diversity of patrilocal societies has been described in tribal populations (Oota et al. 2001), but can also be observed in Europe (Seielstad et al. 1998).

Many of the population structures revealed in the present study could potentially be related to recent historical events. Irrespective of whether these links are true, they would have been less likely to be discernible by Y-SNP analyses, which usually mark much earlier waves of population movement (see the Dutch and Baltic examples above). Single-nucleotide substitutions are known to occur at a rate of approximately $2–5\times10^{-9}$ per generation (Cooper and Krawczak 1993) implying that common selectively neutral SNPs are likely to have arisen before the major expansion of the human population. With their demic diffusion throughout the world, Y-SNPs could therefore only create local genetic structures in populations that had become particularly isolated. With Y-STRs, by contrast, the rapid mutation process constantly generates new genetic variation that allows genetic structures to change more easily. Furthermore, even with their substantial variation in marker-specific mutation rates, Y-STRs usually manifest so many meiotic mutations in evolutionarily related haplotypes that the actual choice of markers is fairly inconsequential to the inferred nature of the genealogical process. In this respect, Y-STRs represent a much more robust population genetic tool than Y-SNPs, which may have substantially different time-depths and can thus create layers of incongruent maps (Jobling and Tyler-Smith 2003).

In summary, we have shown that Y-STRs are capable of resolving male genealogies in Europe to an unparalleled degree. Although it is inherently difficult to prove by Y-STR analysis alone whether a particular genetic division is of recent or prehistoric origin, or whether it would have been detectable by Y-SNPs as well, it suggests that Y-STRs should be considered as the markers of choice for studies of local population structure and recent demographic history. Therefore, sampling in YHRD is currently concentrating upon an expansion of the database to Eurasia as well as to the Americas in order to facilitate investigation of migration processes both to and from Europe that occurred during the past millennium.

## Electronic database information

Y-STR haplotype reference database (YHRD): http://www.yhrd.org; Forensic Laboratory for DNA Research, Leiden: http://www.humgen.nl/fldo/; Geographic Resources Analysis Support System: grass.itc.it; US Land Processes Distributed Active Archive Center: edcdaac.usgs.gov

**Acknowledgements** The authors wish to thank Wulf Schiefenhövel, Andechs, for helpful comments and criticisms, and Brian Fulfrost of the UCSC GIS Technology Lab for valuable advice on geospatial data interpolation and visualisation.

## Publication of this paper is being undertaken on behalf of the Forensic Y-Chromosome Research Group

M. Aler[1], A. Alonso[2], C. Alves[3], M. Alù[4], A. Amorim[3], K. Anslinger[5], E. Arroyo[6], A. Asmundo[7], C. Augustin[8], D. Ballard[9], L. Barbarii[10], G. Bäßler[11], A. Betz[11], G. Bläß[12], E. Bosch[13], W. Branicki[14], A. Brehm[15], M. Brion[16], L. Buscemi[17], L. Caenazzo[18], A. Cagliá[19], E. Carnevali[20], E. Carra[21], A. Carracedo[16], K. Crainic[22], Z. de Battisti[23], D. Dermengiu[24], T. Dobosz[25], R. Dominici[26], B.M. Dupuy[27], A.T. Fernandes[15], I. Furac[28], S. Füredi[29], J. García[2], C. Gehrig[30], M. Gené[31], B. Gornjak Pogorelc[32], B. Glock[33], L. Gusmão[3], M. Hedman[34], K. Hedrich[35], J. Henke [36], L. Henke[36], M. Hidding M[37], C. Hohoff[38], G. Holmlund[39], B. Hoste[40], H.-J Kärgel[35], C. Keyser-Traqui[41], M. Klintschar M[42], S. Kravchenko[43], I. Kremensky[44], T. Kupiec[14], M.V. Lareu[16], B. Legrand[22], E. Liebeherr[12], S.A. Limborska[45], L.A. Livshits[43], A.M. López-Parra[6], M. Lorente[46], B. Ludes [41], P. Martín[2], B. Martinez-Jarreta B[47], M.S. Mesa[6], D.M. Monies[48], M. Nagy[49], P. Nievas[47], S. Noerby[50], M. Nowak[51], K.S. Parreira[16], W. Parson[52], V. Pascali[19], R. Pawlowski[14, 53], A. Piccinini[54], R. Ploski[51], M. Poetsch[55], C. Previderé[56], B. Reichenpfader[57], U. Ricci[58], C. Robino[59], B. Rolf [5], A. Sajantila[34], A. Salas[16], P. Sánchez-Diz[16], U. Schmidt[60], C. Schmitt[37], P.M. Schneider[61], I. Skitsa[62], D. Syndercombe Court[9], R. Szibor[63], A. Tagliabracci[17], J. Teifel-Greding[64], K. Thiele[65], M.L.G Uzielli[58], N. von Wurmb-Schwark[66], R. Wegener[67], M. Wozniak[68], B. Zaharova[44], M.T. Zarrabeitia[69], I. Zupanic Pajnic[32]

[1]Laboratory of Forensic Genetics, Department of Legal Medicine, University of Valencia, Spain; [2]National Institute of Toxicology, Biology Services, Ministry of Justice, Madrid, Spain; [3]Institute of Pathology and Molecular Immunology, University of Porto, Portugal; [4]Department of Morphological and Forensic Sciences, University of Modena and Reggio Emilia, Italy; [5]Institute of Legal Medicine, Ludwig-Maximilians-University, Munich, Germany; [6]Laboratory of Forensic Biology, Universidad Complutense de Madrid, Spain; [7]Institute of Legal Medicine, University of Messina, Italy; [8]Institute of Legal Medicine,

University of Hamburg, Germany; [9]Department of Haematology, Queen Mary School of Medicine and Dentistry, London, UK; [10]National Institute of Legal Medicine, Bucharest, Romania; [11]State Criminal Police Office Baden-Württemberg, Stuttgart, Germany; [12]State Criminal Police Office, Berlin, Germany; [13]Unitat de Biologia Evolutiva, Facultat de Ciències de la Salut i de la Vida, Universitat Pompeu Fabra, Barcelona, Spain; [14]Institute of Forensic Research, Kraków, Poland; [15]Centre of Biological and Geological Sciences, University of Madeira, Funchal, Portugal; [16]Institute of Legal Medicine, University of Santiago de Compostela, Spain; [17]Institute of Legal Medicine, University of Ancona, Italy; [18]Institute of Legal Medicine, University of Padova, Italy; [19]Institute of Legal Medicine, Catholic University, Rome, Italy; [20]Institute of Legal Medicine, University of Perugia, Italy; [21]Department of Cellular Biology and Development, University of Palermo, Italy; [22]Department of Legal Medicine, University René Descartes, Paris, France; [23]Institute of Legal Medicine, University of Verona, Italy; [24]Department of Legal Medicine, University of Medicine and Pharmacy, Bucharest, Romania; [25]Institute of Forensic Medicine, Medical University, Wroclaw, Poland; [26]Section of Legal Medicine, Department of Experimental Biomedicine, University of Pisa, Italy; [27]Institute of Forensic Medicine, University of Oslo, Norway; [28]Department of Forensic Medicine and Criminology, University of Zagreb, Croatia; [29]Hungarian National Police, Budapest, Hungary; [30]Institute of Legal Medicine, University of Geneva, Switzerland; [31]Forensic Genetics Laboratory, University of Barcelona, Spain; [32]Institute of Forensic Medicine, University of Ljubljana, Slovenia; [33]Clinical Department for Blood Group Serology, University of Vienna, Austria; [34]Department of Forensic Medicine, University of Helsinki, Finland; [35]State Criminal Police Office Sachsen-Anhalt, Magdeburg, Germany; [36]Institut für Blutgruppenforschung, Cologne, Germany; [37]Institute of Legal Medicine, University of Cologne, Germany; [38]Institute of Legal Medicine, University of Münster, Germany; [39]The National Board of Forensic Medicine, Department of Forensic Genetics, Linköping, Sweden; [40]National Institute of Criminalistics and Criminology, Brussels, Belgium; [41]Institute of Legal Medicine, University of Strasbourg, France; [42]Institute of Legal Medicine, Martin-Luther-University, Halle-Wittenberg, Germany; [43]Department of Human Genomics, Institute of Molecular Biology and Genetics, Kiev, Ukraine; [44]Laboratory of Molecular Pathology, University Hospital of Obstetrics and Gynaecology, Sofia, Bulgaria; [45]Institute of Molecular Genetics, RAS, Moscow, Russia; [46]Department of Legal Medicine, University of Granada, Spain; [47]Department of Forensic Medicine, University of Zaragoza, Spain; [48]Institute of Forensic Medicine, Medical University, Lublin, Poland; [49]Institute of Legal Medicine, Humboldt-University, Berlin, Germany; [50]Institute of Forensic Medicine, University of Copenhagen, Denmark; [51]Department of Forensic Medicine, Medical Academy Warsaw, Poland; [52]Institute of Legal Medicine, University of Innsbruck, Austria; [53]Institute of Forensic Medicine, Medical University of Gdansk, Poland; [54]Institute of Legal Medicine, University of Milan, Italy; [55]Institute of Legal Medicine, Ernst-Moritz-Arndt-University, Greifswald, Germany; [56]Department of Legal Medicine and Public Health, University of Pavia, Italy; [57]Institute of Legal Medicine, University of Graz, Austria; [58]Centre of Medical Genetics and Molecular Medicine, University of Florence, Italy; [59]Laboratory of Criminalistic Sciences, University of Torino, Italy; [60]Institute of Legal Medicine, Albert-Ludwigs-University, Freiburg, Germany; [61]Institute of Legal Medicine, Johannes-Gutenberg-University, Mainz, Germany; [62]Athens Legal Medicine Department, DNA Analysis Laboratory, Athens, Greece; [63]Institute of Legal Medicine, Otto-von-Guericke-University, Magdeburg, Germany; [64]State Criminal Police Office Bavaria, Munich, Germany; [65]Institute of Legal Medicine, University of Leipzig, Germany; [66]Institute of Legal Medicine, Christian-Albrechts-University, Kiel, Germany; [67]Institute of Legal Medicine, University of Rostock, Germany; [68]Institute of Forensic Medicine, University School of Medical Sciences, Bydgoszcz, Poland; [69]Unit of Legal Medicine, University of Cantabria, Santander, Spain

## References

Ammerman AJ, Cavalli-Sforza LL (1984) The neolithic transition and the genetics of populations in Europe. Princeton University Press, Princeton

Banniard M (1989) Genèse culturelle de l'Europe - Ve au VIII siècle. Le Seuil, Paris

Barbujani G (1987) Autocorrelation of gene frequencies under isolation by distance. Genetics 117:777–782

Barbujani G, Chikhi L (2000) Genetic population structure of Europeans inferred from nuclear and mitochondrial DNA polymorphisms. In: Renfrew C, Boyle K (eds) Archaeogenetics: DNA and the population prehistory of Europe. McDonald Institute for Archaeological Research, Cambridge, pp 119–129

Bartlett R (1994) The making of Europe: conquest, colonization, and cultural change, 950-1350. Princeton University Press, Princeton

Boyd R, Silk JB (1997) How humans evolved. WW Norton, New York

Chikhi L, Nichols RA, Barbujani G, Beaumont MA (2002) Y genetic data support the Neolithic demic diffusion model. Proc Natl Acad Sci U S A 99:11008–11013

Cooper DN, Krawczak M (1993) Human gene mutation. BIOS Scientific Publishers, Oxford

Ewens WJ (1972) The sampling theory of selectively neutral alleles. Theor Popul Biol 3:87–112

Excoffier L, Smouse PE (1994) Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. Genetics 136:343–359

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among mtDNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Gieysztor A, Kieniewicz S, Rostworowski E, Tazbir J, Wereszycki H (1979) History of Poland. Polish Scientific Publishers, Warsaw

Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr W, Morling

N, Olaisen B, Pascali VL, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C (2001) DNA commission of the international society of forensic genetics: recommendations on forensic analysis using Y chromosome STRs. Int J Legal Med 114:305–309

Gimbutas M (1971) The slavs. Thames & Hudson, London

Helgason A, Sigurdadottir S, Nicholson J, Sykes B, Hill EW, Bradley DG, Bosnes V, Gulcher JR, Ward R, Stefansson K (2000) Estimating Scandinavian and gaelic ancestry in the male settlers of Iceland. Am J Hum Genet 67:697–717

Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. Nat Rev Genet 4:598–612

Jobling MA, Hurles M, Tyler-Smith C (2004) Human evolutionary genetics. Garland Science, New York

Kayser M, Cagliá A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Piccinini A, Perez-Lezaun A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, de Knijff P, Roewer L (1997) Evaluation of Y chromosomal STRs: a multicenter study. Int J Leg Med 110:125–133

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M, Sajantila A (2000) 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 66:1580–1588

Kayser M, Krawczak M, Excoffier L, Dieltjes P, Corach D, Pascali V, Gehrig C, Bernini LF, Jespersen J, Bakker E, Roewer L, de Knijff P (2001) Extensive analysis of chromosome Y microsatellite haplotypes from globally dispersed human populations. Am J Hum Genet 68:990–1018

Kittler R, Erler A, Brauer S, Stoneking M, Kayser M (2003) Apparent intrachromosomal exchange on the human Y chromosome explained by population history. Eur J Hum Genet 11:304–314

de Knijff P (2000) Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. Am J Hum Genet 67:1055–1061

Lessig R, Edelmann J, Krawczak M (2001) Population genetics of Y-chromosomal microsatellites in Baltic males. Forensic Sci Int 118:153–157

Lessig R, Edelmann J, Zoledziewska M, Dobosz T, Fahr K, Kostrzewa M (2004) SNP-genotyping on human Y-chromosome for forensic purposes—comparison of two different methods. In: Progress in forensic genetics, vol 10, Elsevier, Amsterdam, pp 334–336

Merriman J (1996) A history of modern Europe. WW Norton, New York

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, Stoneking M (2001) Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nat Genet 29:20–21

Ploski R, Wozniak M, Pawlowski R, Monies DM, Branicki W, Kupiec T, Kloostermann A, Dobosz T, Bosch E, Nowak E, Lessig R, Jobling MA, Roewer L, Kayser M (2002) Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. Hum Genet 110:592–600

Richards M, Macaulay V, Torroni A, Bandelt HJ (2002) In search of geographical patterns in European mitochondrial DNA. Am J Hum Genet 71:1168–1174

Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P (1996) Analysis of molecular variance (AMOVA) of Y-chromosome specific microsatellites in two closely related human populations. Hum Mol Genet 5:1029–1033 (Erratum in Hum Mol Genet 6:828 1997)

Roewer L, Kayser M, de Knijff P, Anslinger K, Cagliá A, Corach D, Füredi S, Henke L, Hidding M, Kärgel H-J, Lessig R, Nagy M, Pascali VL, Parson W, Rolf B, Schmitt C, Szibor R, Teifel-Greding J, Krawczak M (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

Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Cagliá A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kärgel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martínez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M (2001) Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. Forensic Sci Int 118:106–113

Rolf B, Meyer E, Brinkmann B, de Knijff P (1998) Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographic clustering. Eur J Hum Genet 6:583–588

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper GC, Corte-Real H, de Knijff P, Decorte R, Dubrova YE, Evgravov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorovska A, Kalaydijeva L, Kayser M, Kravcenko SA, Lavinha J, Livshits LA, Maria S, McElreavey K, Meitinger TA, Melegh B, Mitchell RJ, Nicholson J, Norby S, Noveletto A, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previderé C, Rajczy K, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Shlumukova M, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA (2000) Y-chromosomal diversity within Europe is clinal and influenced primarily by geography rather than language. Am J Hum Genet 67:1526–1543

Schneider S, Roessli D, Excoffier L (2000) Arlequin: A software for population genetics data analysis, version 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva

Seielstad MT, Minch E, Cavalli-Sforza LL (1998) Genetic evidence for a higher female migration rate in humans. Nat Genet 20:278–180

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, de Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290:1155–1159

Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics134:983–993

Sokal RR, Oden NL (1978) Spatial autocorrelation in biology. 1. Methodology. Biol J Linn Soc 10:199–228

Weale ME, Yepiskoposyan L, Jager RF, Hovhannisyan N, Khudoyan A, Burbage-Hall O, Bradman N, Thomas MG (2001) Armenian Y chromosome haplotypes reveal strong regional structure within a single ethno-national group. Hum Genet 109:659-674

Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG (2002) Y chromosome evidence for Anglo-Saxon mass migration. Mol Biol Evol 19:1008–1021

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. Proc Natl Acad Sci U S A 98:10244–10249