

Y-STR variation among Slavs: evidence for the Slavic homeland in the middle Dnieper basin

Krzysztof Rębała · Alexei I. Mikulich · Iosif S. Tsybovsky ·
Daniela Siváková · Zuzana Džupinková ·
Aneta Szczerkowska-Dobosz · Zofia Szczerkowska

Received: 15 September 2006 / Accepted: 7 February 2007
© The Japan Society of Human Genetics and Springer 2007

Abstract A set of 18 Y-chromosomal microsatellite loci was analysed in 568 males from Poland, Slovakia and three regions of Belarus. The results were compared to data available for 2,937 Y chromosome samples from 20 other Slavic populations. Lack of relationship between linguistic, geographic and historical relations between Slavic populations and Y-short tandem repeat (STR) haplotype distribution was observed. Two genetically distant groups of Slavic populations were revealed: one encompassing all Western-Slavic, Eastern-Slavic, and two Southern-Slavic populations, and one encompassing all remaining Southern

Slavs. An analysis of molecular variance (AMOVA) based on Y-chromosomal STRs showed that the variation observed between the two population groups was 4.3%, and was higher than the level of genetic variance among populations within the groups (1.2%). Homogeneity of northern Slavic paternal lineages in Europe was shown to stretch from the Alps to the upper Volga and involve ethnicities speaking completely different branches of Slavic languages. The central position of the population of Ukraine in the network of insignificant AMOVA comparisons, and the lack of traces of significant contribution of ancient tribes inhabiting present-day Poland to the gene pool of Eastern and Southern Slavs, support hypothesis placing the earliest known homeland of Slavs in the middle Dnieper basin.

K. Rębała (✉) · Z. Szczerkowska
Department of Forensic Medicine,
Medical University of Gdansk,
ul. Marii Skłodowskiej-Curie 3A,
80-210 Gdansk, Poland
e-mail: k.rebala@amg.gda.pl

A. I. Mikulich
Institute for the Study of Arts, Ethnography and Folklore,
National Academy of Sciences, Minsk, Belarus

I. S. Tsybovsky
Institute of Problems of Criminology,
Criminalistics and Forensic Expertise, Minsk, Belarus

D. Siváková
Department of Anthropology, Comenius University,
Bratislava, Slovakia

Z. Džupinková
Department of Experimental and Applied Genetics,
Institute of Preventive and Clinical Medicine,
Slovak Medical University, Bratislava, Slovakia

A. Szczerkowska-Dobosz
Department of Dermatology,
Medical University, Gdansk, Poland

Keywords Y chromosome · Microsatellites ·
Slavic populations

Introduction

Since the human Y chromosome is characterised by the presence of the largest non-recombining region in the whole human genome, sensitivity of genetic variation to drift phenomena, a unique inheritance pattern and specificity to males, its polymorphism has been widely studied by researchers interested in human evolution and forensic geneticists (Jobling and Tyler-Smith 2003; Butler 2003). Depending on the time scales of the population history events, different types of polymorphic markers abundant on the Y chromosome are available for research. Analysis of slowly evolving Y-chromosomal biallelic polymorphisms have enabled deeper insight into prehistoric population movements and colonisation waves in Europe

(Semino et al. 2000; Rootsi et al. 2004). On the other hand, Y-chromosomal short tandem repeats (Y-STRs) are characterised by a relatively high mutation rate and seem to be much more suitable for genetic studies of more-recent events. Although the combination between Y-chromosomal microsatellites and biallelic polymorphisms yields the highest level of resolution and a means of clarifying complex genetic histories (Weale et al. 2001, 2002; Das et al. 2004), Y-STR data alone also provide very useful information for analyses of interpopulation diversity and have been widely applied in resolving differentiation of various human populations (Kayser et al. 2001; Ploski et al. 2002; Caglià et al. 2003; Roewer et al. 2005; Immel et al. 2006).

The Indo-European linguistic family has the largest number of speakers of the recognised families of languages in the world today. In this family, Slavic languages form a group of closely related languages with close to 250 million speakers worldwide (Schenker 1995). In Europe, Slavs are the most numerous ethnic and linguistic group of peoples, residing chiefly in eastern and southeastern Europe but also extending across northern Asia to the Pacific Ocean. The early medieval great migrations in Europe utterly changed the ethnic and linguistic situation of the continent and spread the Slavic settlement in the fifth to sixth centuries over the major part of Eastern Europe, leading to the ethno-cultural subdivision of the primarily united Proto-Slavic community (Encyclopædia Britannica 2006). Nowadays, from the linguistic, cultural and geographic point of view, Slavs are customarily divided into three major subgroups: Eastern Slavs (Belarusians, Russians, Ukrainians), Western Slavs (Poles, Slovaks, Czechs, Lusatians), and Southern Slavs (Slovenes, Croats, Bosnians, Montenegrins, Serbs, Macedonians, Bulgarians).

Since the ethno-cultural subdivision of Slavs and the formation of modern Slavic nations took place quite recently, and significant differences in Y-STR haplotype distribution exist even between closely related human populations (Roewer et al. 1996), hypervariable Y-chromosomal microsatellites seem to be markers of choice for the study of mutual relations between different Slavic ethnic groups. So far, it was demonstrated that Slavic Y-STR haplotype paternal lineages form a separate branch in a phylogenetic tree of European populations (Roewer et al. 2001, 2005). However, no comprehensive analysis of interpopulation Y-STR haplotype variation between different Slavic groups was available. This study has focused on providing a phylogenetic overview of closely related Central-Eastern European populations of Poland, Slovakia and Belarus to analyse their relationship with each other and with other Slavic populations, and to investigate how these relations reflect Slavs' historical migrations.

Materials and methods

Eighteen Y-chromosomal microsatellite loci: DYS19, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426, DYS437, DYS438, DYS439, DYS460, GATA H4.1, DYS385 a/b, and YCAII a/b, were genotyped in randomly selected, unrelated Poles ($n = 208$), Slovaks ($n = 164$) and Belarusians ($n = 196$) by means of a multiplex polymerase chain reaction (PCR) and capillary electrophoresis, as previously described (Rębała and Szczerkowska 2005). The Belarusian population included samples from three distinct regions: 86 males from southern Belarus, 57 males from central Belarus, and 53 males from northern Belarus. Additionally, all Belarusian samples were genotyped at the M46 (Tat) locus by a PCR restriction fragment-length polymorphism (RFLP) method (Kayser et al. 2005). The products of amplification of the M46 locus were digested with Hsp92II restriction endonuclease (Rybakowski et al. 2002), separated by polyacrylamide gel electrophoresis, and visualised by silver staining. The Y-STR haplotype data for other 20 Slavic ($n = 2,937$) and nine neighbouring non-Slavic ($n = 1,428$) populations were obtained either from the Y chromosome haplotype reference database (<http://www.ystr.org>) (Roewer et al. 2001) or from the literature (Pepinski et al. 2004a, b; Lauc et al. 2005; Lovrečić et al. 2005; Marjanovic et al. 2005; Peričić et al. 2005; Spiroski et al. 2005; Lessig et al. 2006; Rodig et al. 2007).

Analysis of molecular variance (AMOVA) was performed with the use of Arlequin 3.1 software (Excoffier et al. 2005) to calculate matrices of pairwise F_{ST} and R_{ST} values between populations. Associated probability values were estimated from 10,100 permutations. Linearised F_{ST} and R_{ST} values (Slatkin 1995) were applied to build a neighbour-joining tree using the options NEIGHBOUR and DRAWTREE in the PHYLIP package (Felsenstein 2004), and to conduct a multidimensional scaling analysis using the STATISTICA 7.1 software (StatSoft). In all calculations, DYS389 was considered as a haplotype of two independent loci: DYS389I (repetitive stretches: $p + q$) and DYS389II-I (repetitive stretches: $m + n$) (Rolf et al. 1998), while microvariants, null alleles and locus multiplications were treated as missing data.

Results and discussion

AMOVA of 18-locus Y chromosome haplotypes revealed significant differences between the populations of Belarus, Poland and Slovakia ($F_{ST} = 0.0100$; $P < 0.0001$) as well as between the three Belarusian subpopulations ($F_{ST} = 0.0084$; $P = 0.0034$). Analysis of pairwise F_{ST} values for 18 Y-STRs demonstrated that heterogeneity within the

Belarusian population was caused by differences between northern and central Belarus while the remaining pairwise comparisons did not yield statistically significant F_{ST} values. The most outstanding populations were those of Poland and northern Belarus, while populations of central Belarus, southern Belarus and Slovakia were genetically indistinguishable. The Polish population was shown to be distinct from all studied populations ($P = 0.01$ for comparison with Slovakia, $P < 0.005$ for all other comparisons), while in the case of northern Belarus, the only insignificant F_{ST} value was for a comparison with the southern part of the country.

Our results have demonstrated that the Y-STR haplotype distribution does not reflect the linguistic and/or historical affiliations between Slavic populations. The Polish paternal lineages revealed by Y-chromosomal microsatellite haplotype analysis were previously reported to be distant from a number of non-Slavic European populations and Slavic-speaking Muscovites (Ploski et al. 2002), but no Slavic populations inhabiting Poland's geographic neighbours were taken into consideration. Despite the usage of very closely related languages by Poles and Slovaks as well as over half a millennium of a common state bordering the majority of Poles and Belarusians, the Polish Y-STR haplotype heritage was shown to be distinct from that of both neighbouring Slavic nations. On the other hand, analysis of Y chromosome haplotypes defined by 18 loci has revealed genetic homogeneity between Slovaks and two subpopulations of Belarus, although both populations are geographically distant and isolated, speak tongues belonging to separate branches of the Slavic language group, and have never shared common state borders throughout their histories.

Since for the majority of other Slavic populations, only nine-locus haplotypes (DYS19, DYS389I, DYS389II-I, DYS390, DYS391, DYS392, DYS393, and DYS385 a/b) were available, only these loci were used for further analysis. A comparison of our Polish haplotypes defined by the selected nine loci with data for six Polish regional subpopulations from the Y chromosome haplotype reference database confirmed previously observed homogeneity of Polish subpopulations ($F_{ST} = -0.0003$; $P = 0.66$) (Ploski et al. 2002), and in AMOVA, all Polish samples were combined. Results revealed four clusters of Slavic populations connected by a network of statistically insignificant F_{ST} values ($P > 0.05$): (1) all Western-Slavic and Eastern-Slavic populations, Slovenes and western Croats; (2) Lusatians; (3) Southern-Slavic northern Croats and Bosnians; (4) Southern-Slavic Serbs, Macedonians, and Bulgarians (Table 1). However, at the significance level of 0.01, only one such cluster involving all Slavic populations was disclosed. The distinctiveness of Southern-Slavic populations was observed as a separate branch in a neighbour-joining

tree, while multidimensional scaling has displayed a nucleus of seven genetically indistinguishable populations with very small relative genetic distances, which involved population samples of Eastern-Slavic (Ukrainian, Russian, Belarusian), Western-Slavic (Slovak) as well as Southern-Slavic (Slovene) origin (Fig. 1).

Since there is no clear consensus over the accuracy of different statistical parameters estimating genetic distances between populations in studies using microsatellite markers, both a classical allele frequency-based differentiation estimator (F_{ST}) and its stepwise mutation model-based analogue (R_{ST}) are commonly reported (Balloux and Lugin-Moulin 2002). Therefore, we applied both distance methods to assess genetic relations between various Slavic and non-Slavic populations. In accordance with results obtained for autosomal STRs in various human populations (Pérez-Lezaun et al. 1997; Destro-Bisol et al. 2000) and for Y-chromosomal microsatellites in sub-Saharan Africans (Caglià et al. 2003), the pattern of Y-STR interpopulation diversity among Slavs and neighbouring populations, based on F_{ST} values, appeared to be congruent with known ethno-historical relationships, while that based on R_{ST} values revealed unexpected and unconvincing population affinities.

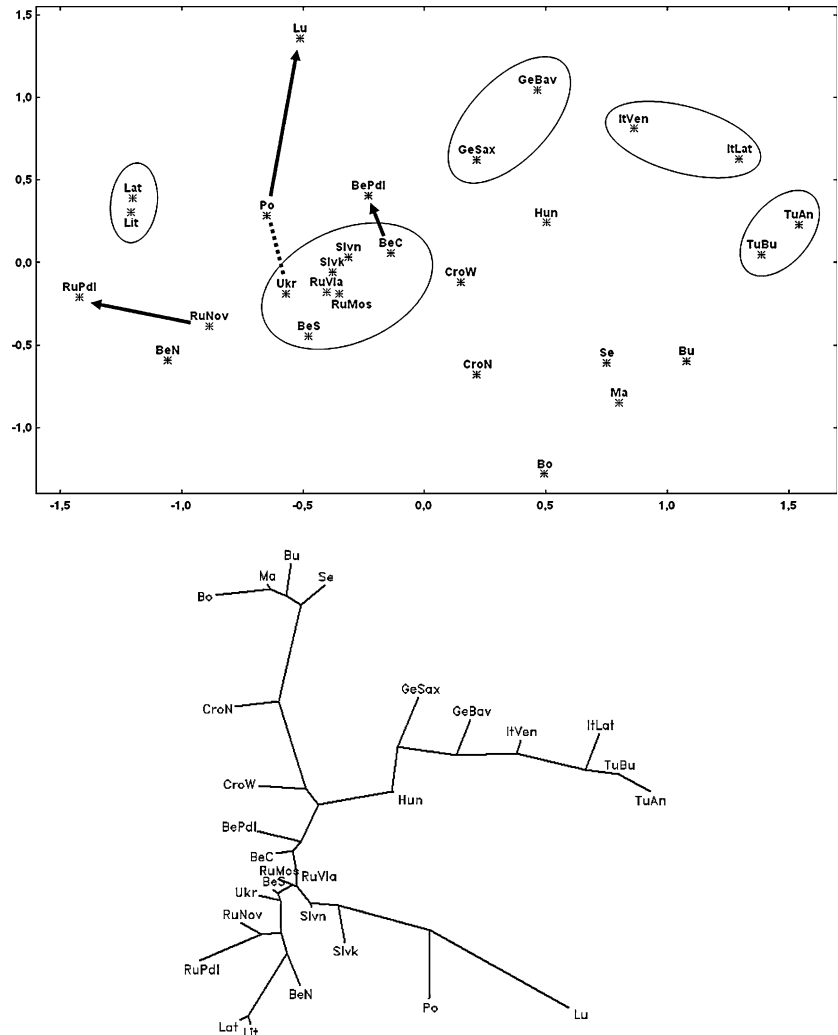
Both the multidimensional scaling plot and the neighbour-joining tree, based on the F_{ST} values, revealed genetic proximity between related populations: (1) Germans from Bavaria and Saxony, (2) Italians from Latium and Veneto, (3) Turks from Anatolia and Bulgaria, and (4) Balts from Latvia and Lithuania (Fig. 1). The F_{ST} -based results were consistent with expectations also in case of three isolated Slavic populations: (1) Lusatians from southeastern Germany, who are descendants of Slavic tribes that have inhabited the lands between the Elbe and Oder rivers since the fifth century (Encyclopædia Britannica 2006), (2) Polish Belarusians, who colonised parts of Podlachia (northeastern Poland) in the 15th to 16th centuries after arrival from the Hrodna region (Wiśniewski 1964), and (3) a community of Russian settlers (Old Believers), who arrived in Podlachia in the eighteenth century from the Pskov and Novgorod regions (Grek-Pabisowa 1968). All three ethnic groups were shown to be homogeneous with only one of all compared populations (Table 1), representing their population of origin. Lusatians revealed Y-STR homogeneity with the neighbouring population inhabiting areas from which they are supposed to migrate, i.e. with Poles. In the case of Podlachian Belarusians, such a population of origin was geographically the closest population of central Belarus (involving the city of Hrodna), while in the case of Polish Old Believers it was Russians from the Novgorod region.

On the other hand, in the R_{ST} -based multidimensional scaling plot and the neighbour-joining tree, only the two

Table 1 Y-STR haplotype pairwise F_{ST} values below the diagonal and corresponding P values above the diagonal (P values > 0.05 are bold) for 19 Slavic populations

	Lu <i>n</i> = 29	Po <i>n</i> = 1521	Slvk <i>n</i> = 164	BeN <i>n</i> = 53	BeC <i>n</i> = 57	BeS <i>n</i> = 86	BePdI <i>n</i> = 157	RuPdI <i>n</i> = 127	RuNov <i>n</i> = 50	RuMos <i>n</i> = 85	RuVla <i>n</i> = 152	Ukr <i>n</i> = 82	Slvn <i>n</i> = 121	CroW <i>n</i> = 101	CroN <i>n</i> = 150	Bo <i>n</i> = 100	Se <i>n</i> = 114	Ma <i>n</i> = 234	Bu <i>n</i> = 122			
Lu	-	0.0336	0.0036	0.0000	0.0003	0.0000	0.0001	0.0000	0.0003	0.0001	0.0001	0.0023	0.0013	0.0027	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000		
Po	0.0139	-	0.0186	0.0000	0.0003	0.0000	0.0000	0.0000	0.0055	0.0050	0.0000	0.0557	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Slvk	0.0280	0.0032	-	0.0026	0.0884	0.1174	0.0000	0.0000	0.0530	0.5154	0.0168	0.8820	0.6930	0.0544	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
BeN	0.0962	0.0326	0.0185	-	0.0129	0.1243	0.0001	0.0048	0.2636	0.0372	0.0126	0.2224	0.0006	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
BeC	0.0543	0.0188	0.0062	0.0215	-	0.1404	0.0564	0.0012	0.0554	0.3186	0.4992	0.1489	0.2622	0.0558	0.0013	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
BeS	0.0558	0.0181	0.0040	0.0068	0.0059	-	0.0003	0.0006	0.2481	0.3949	0.0431	0.6576	0.0352	0.0138	0.0062	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	
BePdI	0.0453	0.0266	0.0144	0.0300	0.0075	0.0171	-	0.0000	0.0007	0.0009	0.0004	0.0012	0.0000	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
RuPdI	0.0623	0.0232	0.0216	0.0188	0.0230	0.0210	0.0348	-	0.1002	0.0002	0.0000	0.0039	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
RuNov	0.0562	0.0134	0.0087	0.0036	0.0127	0.0034	0.0243	0.0068	-	0.1632	0.1102	0.2940	0.0279	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
RuMos	0.0479	0.0084	-0.0006	0.0132	0.0017	0.0003	0.0157	0.0201	0.0055	-	0.7386	0.5456	0.3909	0.1102	0.0009	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
RuVla	0.0501	0.0144	0.0064	0.0140	-0.0008	0.0066	0.0132	0.0247	0.0061	-0.0023	-	0.0960	0.0181	0.0045	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Ukr	0.0392	0.0042	-0.0037	0.0038	0.0054	-0.0024	0.0164	0.0143	0.0023	-0.0012	0.0045	-	0.5272	0.0246	0.0008	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
Slvn	0.0379	0.0073	-0.0016	0.0245	0.0024	0.0078	0.0188	0.0271	0.0120	0.0004	0.0069	-0.0009	-	0.1419	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	
CroW	0.0336	0.0166	0.0053	0.0424	0.0091	0.0115	0.0158	0.0516	0.0258	0.0050	0.0109	0.0098	0.0035	-	0.0396	0.0001	0.0032	0.0000	0.0000	0.0001	0.0001	
CroN	0.0583	0.0357	0.0154	0.0413	0.0219	0.0119	0.0316	0.0703	0.0378	0.0171	0.0232	0.0193	0.0178	0.0064	-	0.0975	0.0207	0.0034	0.0000	0.0000	0.0000	
Bo	0.0946	0.0681	0.0434	0.0680	0.0546	0.0297	0.0595	0.1051	0.0646	0.0435	0.0526	0.0483	0.0525	0.0268	0.0044	-	0.0161	0.0208	0.0000	0.0000	0.0000	
Se	0.0642	0.0541	0.0351	0.0744	0.0326	0.0353	0.0441	0.0956	0.0610	0.0328	0.0392	0.0439	0.0397	0.0132	0.0075	0.0103	-	0.4738	0.2041	0.0891	0.0891	
Ma	0.0825	0.0638	0.0430	0.0751	0.0420	0.0420	0.0530	0.1077	0.0657	0.0401	0.0447	0.0509	0.0468	0.0191	0.0083	0.0074	-0.0003	-	0.0891	0.0891	0.0891	
Bu	0.0935	0.0738	0.0509	0.0879	0.0422	0.0508	0.0542	0.1132	0.0803	0.0430	0.0516	0.0605	0.0537	0.0220	0.0205	0.0197	0.0023	0.0031	-	-	-	-

Fig. 1 Two-dimensional plot obtained from multidimensional scaling, and a neighbour-joining tree, based on pairwise F_{ST} values for nine-locus Y-STR haplotypes observed in 19 Slavic and 9 non-Slavic populations. *Ellipses* are traced around genetically homogeneous northern Slavic populations and clusters of non-Slavic populations with known ethno-historical affiliations. *Arrows* indicate historically proved directions of migrations. A *dotted line* connects populations with a disputed direction of migration, which inhabit areas designated according to the various sources as the Slavic homeland



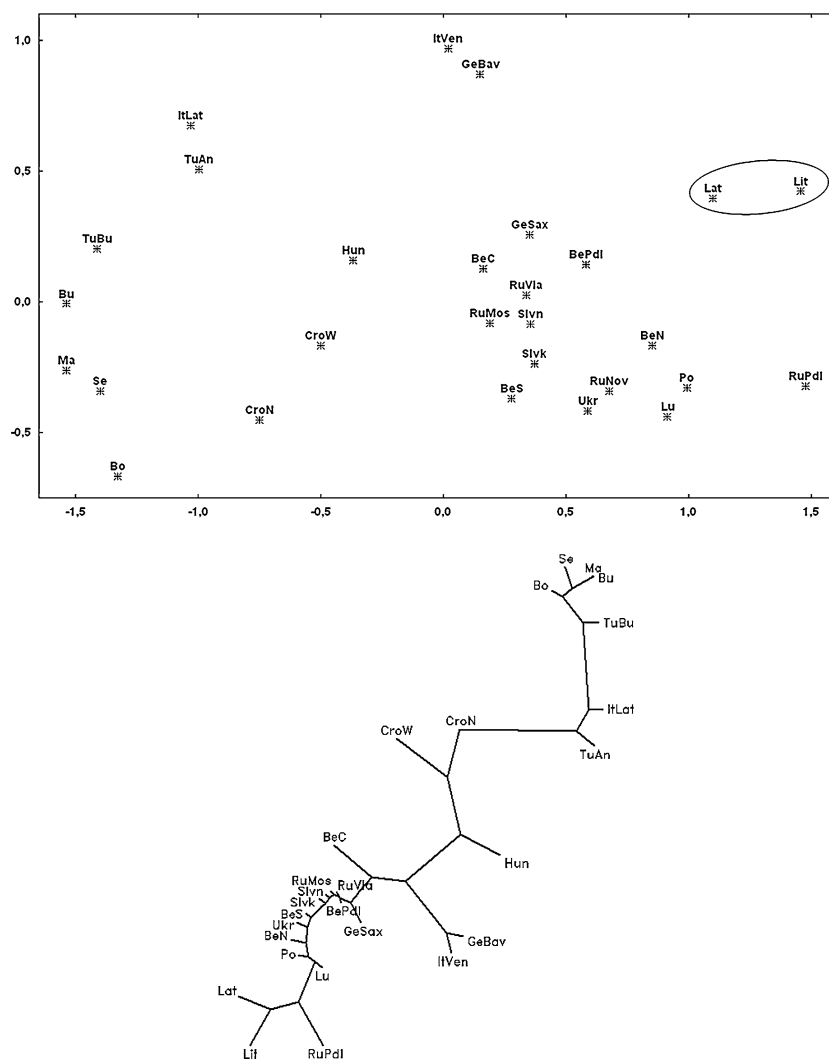
Baltic populations were clearly isolated from unrelated ethnicities, while separation of German, Italian and Turkish populations was no longer visible (Fig. 2). Moreover, statistically insignificant P values for R_{ST} genetic distances were observed between historically, linguistically, and geographically unrelated populations such as Turks and Italians from Latium, or Belarusians and Germans from Saxony. In the case of the three isolated Slavic populations, genetic homogeneity restricted only to the populations of origin, as observed for F_{ST} values, was lost in the R_{ST} method. Thus, the F_{ST} genetic distances reflect interpopulation relationships between the compared populations much better than their stepwise-based analogues. This indicates that the genetic differences in the Y-STR haplotype distribution between the Slavic populations and their nearest neighbours are caused mainly by drift, and is explained by the fact that a period of time since the differentiation of the compared populations was too short to allow for an effective impact of mutations on Y-chromosome variation (Pérez-Lezaun et al. 1997; Caglià et al.

2003). Therefore, in further analysis, we focused only on the results obtained by the classical allele frequency-based approach for estimation of Y-STR genetic distances.

Comprehensive analysis of Slavic Y-chromosomal microsatellite haplotypes on a European scale confirmed previous observations for 18 Y-STR loci in the Polish, Slovak and Belarusian populations: no relationship between the customary linguistic division of Slavs and the Y-STR haplotype distribution was disclosed. The most apparent genetic distance was found between the northern (Eastern and Western) and Southern Slavs, who at the end of the 9th century were separated by the invasion of Finno-Ugric Hungarians. The AMOVA showed that the variation observed between both population groups was 4.3% ($F_{CT} = 0.0428$; $P = 0.0003$), which was higher than the level of genetic variance among populations within the groups (1.2%; $F_{SC} = 0.0130$; $P < 0.0001$). This difference was even more profound when the R_{ST} -based distance method was applied: genetic variation between both population clusters was 19.8% ($R_{CT} = 0.1984$; $P = 0.0002$)

Fig. 2 Two-dimensional plot obtained from multidimensional scaling, and a neighbour-joining tree, based on pairwise R_{ST} values for nine-locus Y-STR haplotypes observed in 19 Slavic and nine non-Slavic populations. An ellipse is traced around a cluster of populations with known ethno-historical affiliations

Lu Lusatians; *Po* Poles; *Slvk* Slovaks; *BeN*, *BeC*, *BeS*, *BePdl* Belarusians from northern Belarus, central Belarus, southern Belarus, Podlachia; *RuPdl*, *RuNov*, *RuMos*, *RuVla* Russians from Podlachia (Old Believers), Novgorod region, Moscow region, Vladivostok region; *Ukr* Ukrainians; *Slvn* Slovenes; *CroW*, *CroN* Croats from western Croatia, northern Croatia (Zagreb region); *Bo* Bosnians; *Se* Serbs; *Ma* Macedonians; *Bu* Bulgarians; *Lit* Lithuanians; *Lat* Latvians; *GeBav*, *GeSax* Germans from Bavaria, Saxony (Dresden region); *ItLat*, *ItVen* Italians from Latium, Veneto; *Hun* Hungarians; *TuAn*, *TuBu* Turks from Anatolia, Bulgaria



while the interpopulation variance within the groups was only 1.8% ($R_{SC} = 0.0228$; $P < 0.0001$). The observed northern Slavic Y-STR genetic homogeneity extends from Slovakia and Ukraine to parts of Russia and Belarus, but also involves Southern-Slavic populations of Slovenia and western Croatia, and is the most probably due to a homogeneous genetic substrate inherited from the ancestral Slavic population. However, due to the Y-STR proximity of linguistically and geographically Southern-Slavic Slovenes and western Croats to the northern Slavic branch, the observed genetic differentiation cannot simply be explained by the separation of both Slavic-speaking groups by the non-Slavic Romanians, Hungarians, and German-speaking Austrians. A similar difference has been previously reported between Bulgarians and a few other Slavic populations (Roewer et al. 2005), and our results demonstrate that other Southern-Slavic populations, namely Macedonians, Serbs, Bosnians, and northern Croats are genetically distinct from their northern linguistic relatives

as well. Roewer et al. (2005) attributed a possible explanation for these differences to the admixture of Y chromosomes of Finno-Ugric and Turkic-speaking peoples who had invaded and settled in the Danube basin and the Balkans. However, we found that the only population that revealed an insignificant F_{ST} value in comparison with the Finno-Ugric Hungarians was the population of western Croatia, and this putative admixture did not significantly affect the Y-STR proximity of western Croats to Eastern and Western Slavs (Table 1). All other F_{ST} values obtained for comparison of nine-locus haplotypes of Hungarians and Turks from Bulgaria and Anatolia, with those of 19 Slavic populations appeared to be statistically significant ($P = 0.01$ for a comparison of Hungarians with the population of central Belarus, $P = 0.002$ for a comparison of Bulgarian Turks with Bulgarians, $P \leq 0.0003$ for all other comparisons). Thus, the contribution of the Y chromosomes of peoples who settled in the region before the Slavic expansion to the genetic heritage of Southern Slavs

is the most likely explanation for this phenomenon. On the other hand, our results indicate no significant genetic traces of pre-sixth-century inhabitants of present-day Slovenia in the Slovene Y chromosome genetic pool.

Although the existence of the Balto-Slavic linguistic community, or at least territorial contiguity of Proto-Baltic and Proto-Slavic in the past, is generally accepted (Schenker 1995), AMOVA revealed significant differences in Y-STR distribution between Slavic and Baltic populations ($P < 0.005$ for all pairwise comparisons), which is likely to result from the previously observed different Y-chromosomal haplogroup distribution (Rosser et al. 2000). The Baltic populations are characterised by the high incidence of the Y-chromosomal haplogroup N3 (47% among Lithuanians, 32% among Latvians) (Rosser et al. 2000; Zerjal et al. 2001). Its distribution pattern in Slavic populations indicates that Proto-Slavs did not carry this lineage at a substantial frequency, since it is relatively rare among Slavs and at high frequency was observed only in some Russian subpopulations (Malyarchuk et al. 2004). Since at the significance level equal to 0.01, the only population that did not yield statistically significant F_{ST} values for comparisons with other Slavic and both Baltic populations was the Slavic-speaking population of northern Belarus, we estimated haplogroup N3 frequencies in the three Belarusian subpopulations. The results suggest that the uniqueness of the northern Belarusian population is most likely due to the high incidence of Y chromosomes from the haplogroup N3 (18.9%), which has half the frequency in central and southern Belarus (8.8 and 8.1%, respectively). Therefore, although the early ethnogenesis of the Belarusian nation has customarily been linked to the gradual Slavicisation of the homogeneous Baltic substrate on the territory of present-day Belarus (Sedov 1970), only northern Belarus seems to be a transient area for the Baltic and Slavic settlement. Apart from Balts, the N3 Y chromosomes are also prevalent among Finno-Ugrians (Rosser et al. 2000; Zerjal et al. 2001), but it was found that Y-STR haplotypes from the haplogroup N3 differ in Baltic and Finno-Ugric populations, most likely due to two distinct migration waves of people carrying N3 Y chromosomes (Zerjal et al. 2001; Kasperavičiūtė et al. 2004). On the contrary to the Finno-Ugric populations, where allele DYS19*14 is the most common among haplogroup N3 males, allele DYS19*15 is the most frequent in the Baltic haplogroup N3 (Kasperavičiūtė et al. 2004), and was also the most frequent allele in the haplogroup N3 in northern Belarus (60.0%), providing additional evidence for the presence of the Baltic substrate in the genetic pool of Belarusians from the northernmost part of the country.

Localisation of the Slavic homeland prior to their great expansion in the fifth to sixth centuries is one of the key

problems of European history in the first millennium AD. Although it is assumed that prehistorically the original habitat of Slavs was Asia, from which they migrated in the third or second millennium BC to populate parts of Eastern Europe (Encyclopædia Britannica 2006), a debate concerning the European homeland of Slavs seems to remain unsolved. Because Slavs unequivocally enter the records of history as late as the sixth century AD, when their expansion in Eastern Europe was already advanced, different theories concerning the Slavs' geographic origin based on archaeological, anthropological and/or linguistic data have been formulated. Two such theories have gained the largest support among the scientists (Schenker 1995), one placing the cradle of Slavs in the watershed of the Vistula and Oder rivers (present-day Poland), and the other locating it in the watershed of the middle Dnieper (present-day Ukraine). Our results indicate that using the population-of-origin approach based on the AMOVA, as many as nine ($P > 0.05$) or ten ($P > 0.01$) populations can be traced back to the lands of present-day Ukraine, including Eastern-Slavic Russians and Belarusians, Western-Slavic Poles and Slovaks, and Southern-Slavic Slovenes and Croats. On the other hand, the Polish population gave insignificant F_{ST} values in pairwise comparisons with only one (i.e. Ukrainians) or three (i.e. Ukrainians, Slovaks, and Lusatians) populations ($P > 0.05$ or 0.01, respectively). Moreover, the Y-STR genetic distance between Poles and Belarusians, who are geographic neighbours (Table 1), excludes significant gene flow between the two populations and localisation of Belarusians' ancestors in present-day Poland.

In conclusion, we have demonstrated that Y-STR haplotype distribution divides Slavs into two genetically distant groups: one encompassing all Western Slavs, Eastern Slavs, Slovenes and Western Croats, and the other involving all remaining Southern Slavs. Many northern Slavic populations are genetically indistinguishable in regard to the nine-locus Y-STR haplotype variation, and this homogeneity extends from the Alps to the upper Volga, and even as far as the Pacific Ocean (eastern Russia), regardless of linguistic, cultural and historical affiliations of the Slavic ethnicities. The example of Slovaks and Belarusians shows that this homogeneity is likely to be extended to other Y-chromosomal microsatellites as well. Results of the interpopulation Y-STR haplotype analysis exclude a significant contribution of ancient tribes inhabiting present-day Poland to the gene pool of Eastern and Southern Slavs, and suggest that the Slavic expansion started from present-day Ukraine, thus supporting the hypothesis that places the earliest known homeland of Slavs in the basin of the middle Dnieper. To our knowledge, this is the first report on the use of genetic markers in solving the question of the localisation of the Slavic homeland.

References

- Balloux F, Lugon-Moulin N (2002) The estimation of population differentiation with microsatellite markers. *Mol Ecol* 11:155–165
- Butler JM (2003) Recent developments in Y-short tandem repeat and Y-single nucleotide polymorphism analysis. *Forensic Sci Rev* 15:91–111
- Caglià A, Tofanelli S, Coia V, Boschi I, Pescarmona M, Spedini G, Pascali V, Paoli G, Destro-Bisol G (2003) A study of Y-chromosome microsatellite variation in sub-Saharan Africa: a comparison between F_{ST} and R_{ST} genetic distances. *Hum Biol* 75:313–330
- Das B, Chauhan PS, Seshadri M (2004) Minimal sharing of Y-chromosome STR haplotypes among five endogamous population groups from western and southwestern India. *Hum Biol* 76:743–763
- Destro-Bisol G, Spedini G, Pascali VL (2000) Application of different genetic distance methods to microsatellite data. *Hum Genet* 106:130–132
- Encyclopædia Britannica (2006) Slav in Encyclopædia Britannica Online (<http://www.search.eb.com/eb/article-9068162>)
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47–50
- Felsenstein J (2004) PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle
- Grek-Pabisowa I (1968) Rosyjska gwara starowierców w województwach olsztyńskim i białostockim. Russian dialect of Old-Believers in the administrative regions of Olsztyn and Białystok. Ossolineum, Wrocław
- Immel U-D, Krawczak M, Udolph J, Richter A, Rodig H, Kleiber M, Klintschar M (2006) Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population. *Eur J Hum Genet* 14:577–582
- Jobling MA, Tyler-Smith C (2003) The human Y chromosome: an evolutionary marker comes of age. *Nat Rev Genet* 4:598–612
- Kasperavičiūtė D, Kučinskas V, Stoneking M (2004) Y chromosome and mitochondrial DNA variation in Lithuanians. *Ann Hum Genet* 68:438–452
- 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) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 68:990–1018
- Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L, Hohoff C, Illing A, Jonkisz A, Kuzniar P, Lebioda A, Lessig R, Lewicki S, Maciejewska A, Monies DM, Pawłowski R, Poetsch M, Schmid D, Schmidt U, Schneider PM, Stradmann-Bellinghausen B, Szibor R, Wegener R, Wozniak M, Zoledziewska M, Roewer L, Dobosz T, Ploski R (2005) Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet* 117:428–443
- Lauc LB, Peričić M, Klarić IM, Šijački A, Popović D, Janičijević B, Rudan P (2005) Y chromosome STR polymorphisms in a Serbian population sample. *Forensic Sci Int* 150:97–101
- Lessig R, Edelmann J, Kleemann WJ, Kozhemyako V (2006) Population data of Y-chromosomal STRs in Russian males of the Primorye region population. *Forensic Sci Int* 159:71–76
- Lovrečić L, Ristić S, Brajenović B, Kapović M, Peterlin B (2005) Human Y-specific STR haplotypes in the Western Croatian population sample. *Forensic Sci Int* 149:257–261
- Malyarchuk B, Derenko M, Grzybowski T, Lunkina A, Czarny J, Rychkov S, Morozova I, Denisova G, Miściska-Śliwka D (2004) Differentiation of mitochondrial DNA and Y chromosomes in Russian populations. *Hum Biol* 76:877–900
- Marjanovic D, Bakal N, Pojskic N, Kapur L, Drobnic K, Primorac D, Bajrovic K, Hadziselimovic R (2005) Population data for the twelve Y-chromosome short tandem repeat loci from the sample of multinational population in Bosnia and Herzegovina. *J Forensic Sci* 50:223–225
- Pepinski W, Niemcunowicz-Janica A, Skawronska M, Koc-Zorawska E, Janica J, Soltyszewski I (2004a) Y-chromosome STR haplotypes in a population sample of the Byelorussian minority living in the northeastern Poland. *Forensic Sci Int* 140:117–121
- Pepinski W, Niemcunowicz-Janica A, Skawronska M, Koc-Zorawska E, Janica J, Soltyszewski I (2004b) Y-chromosome STR haplotypes and alleles in the population sample of Old Believers residing in the Northeastern Poland. *Forensic Sci Int* 143:65–68
- Pérez-Lezaun A, Calafell F, Mateu E, Comas D, Ruiz-Pacheco R, Bertranpetit J (1997) Microsatellite variation and the differentiation of modern humans. *Hum Genet* 99:1–7
- Peričić M, Klarić IM, Lauc LB, Janičijević B, Dorđević D, Efremska L, Rudan P (2005) Population genetics of 8 Y chromosome STR loci in Macedonians and Macedonian Romani (Gypsy). *Forensic Sci Int* 154:257–261
- Ploski R, Wozniak M, Pawłowski R, Monies DM, Branicki W, Kupiec T, Kloosterman A, Dobosz T, Bosch E, Nowak M, 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
- Rębała K, Szczerkowska Z (2005) Polish population study on Y chromosome haplotypes defined by 18 STR loci. *Int J Legal Med* 119:303–305
- Rodig H, Grum M, Grimmecke H-D (2007) Population study and evaluation of 20 Y-chromosome STR loci in Germans. *Int J Legal Med* 121:24–27
- 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
- Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Caglià A, Carracedo A, Corach D, Dekairelle A-F, Dobosz T, Dupuy BM, Füredi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kärger HJ, de Knijff P, Lessig R, Liebherr E, Lorente M, Martínez-Jarreta B, Nieves 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
- Roewer L, Croucher PJP, Willuweit S, Lu TT, Kayser M, Lessig R, de Knijff P, Jobling MA, Tyler-Smith C, Krawczak M (2005) Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 116:279–291
- 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
- Roots S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Peričić M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan R, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan J-P, Chaventre A, Ferak V, Füredi

- S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzić R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. *Am J Hum Genet* 75:128–137
- Rosser ZH, Zerjal T, Hurler ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckman G, Beckman L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper G, Corte-Real HB, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Golge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kivisild T, Kravchenko SA, Krumina A, Kucinskas V, Lavinha J, Livshits LA, Malaspina P, Maria S, McElreavey K, Meitinger TA, Mikelsaar AV, Mitchell RJ, Nafa K, Nicholson J, Norby S, Pandya A, Parik J, Patsalis PC, Pereira L, Peterlin B, Pielberg G, Prata MJ, Previdere C, Roewer L, Rootsi S, Rubinsztein DC, Saillard J, Santos FR, Stefanescu G, Sykes BC, Tolun A, Villems R, Tyler-Smith C, Jobling MA (2000) Y-chromosomal diversity in Europe is clinal and influenced primarily by geography rather than by language. *Am J Hum Genet* 67:1526–1543
- Rybakowski JK, Borkowska A, Czerski PM, Hauser J (2002) Eye movement disturbances in schizophrenia and a polymorphism of catechol-*O*-methyltransferase gene. *Psychiatry Res* 113:49–57
- Schenker AM (1995) *The dawn of Slavic: an introduction to Slavic philology*. Yale University Press, New Haven
- Sedov VV (1970) Slavâne Verhnego Podneprov'â i Podvin'â. The Slavs of the basin of the upper Dnieper and Western Dvina. Nauka, Moscow
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikie 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
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Spiroski M, Arsov T, Krüger C, Willuweit S, Roewer L (2005) Y-chromosomal STR haplotypes in Macedonian population samples. *Forensic Sci Int* 148:69–73
- 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
- Wiśniewski J (1964) Rozwój osadnictwa na pograniczu polsko-rusko-litewskim od końca XIV do połowy XVII wieku. Development of the settlement in the Polish–Ruthenian–Lithuanian borderland from the end of the 14th century to the middle of the 17th century. *Acta Balt Slav* 1:115–136
- Zerjal T, Beckman L, Beckman G, Mikelsaar A-V, Krumina A, Kučinskas V, Hurler ME, Tyler-Smith C (2001) Geographical, linguistic and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations. *Mol Biol Evol* 18:1077–1087