

Rafal Ploski · Marcin Wozniak · Ryszard Pawlowski
Dorota Marta Monies · Wojciech Branicki
Tomasz Kupiec · Ate Kloosterman · Tadeusz Dobosz
Elena Bosch · Magdalena Nowak · Rüdiger Lessig
Mark A. Jobling · Lutz Roewer · Manfred Kayser

Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis

Received: 8 January 2002 / Accepted: 5 March 2002 / Published online: 22 May 2002

© Springer-Verlag 2002

Abstract Different regional populations from Poland were studied in order to assess the genetic heterogeneity within Poland, investigate the genetic relationships with

Electronic supplementary material to this paper can be obtained by using the Springer LINK server located at <http://dx.doi.org/10.1007/s00439-002-0728-0>.

R. Ploski · M. Nowak
Human Molecular Genetics Lab, Department of Forensic Medicine,
Warsaw Medical Academy, Poland

M. Wozniak
Forensic Medicine Institute,
L. Rydygier's University School of Medical Sciences,
Bydgoszcz, Poland

R. Pawlowski
Institute of Forensic Medicine, Medical University Gdańsk,
Poland

D.M. Monies
Institute of Forensic Medicine, Medical University Lublin,
Poland

W. Branicki · T. Kupiec
Institute of Forensic Research, Krakow, Poland

A. Kloosterman
Netherlands Forensic Institute, Rijswijk, The Netherlands

T. Dobosz
Institute of Forensic Medicine, Medical University Wrocław,
Poland

E. Bosch · M.A. Jobling
Department of Genetics, University of Leicester,
United Kingdom

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

L. Roewer
Institute for Legal Medicine, Humboldt-University Berlin,
Germany

M. Kayser (✉)
Department for Evolutionary Genetics,
Max Planck Institute for Evolutionary Anthropology,
Inselstrasse 22, 04103 Leipzig, Germany
e-mail: kayser@eva.mpg.de,
Tel.: +49-341-9952506, Fax: +49-341-9952555

other European populations and provide a population-specific reference database for anthropological and forensic studies. Nine Y-chromosomal microsatellites were analysed in a total of 919 unrelated males from six regions of Poland and in 1,273 male individuals from nine other European populations. AMOVA revealed that all of the molecular variation in the Polish dataset is due to variation within populations, and no variation was detected among populations of different regions of Poland. However, in the non-Polish European dataset 9.3% ($P < 0.0001$) of the total variation was due to differences among populations. Consequently, differences in R_{ST} -values between all possible pairs of Polish populations were not statistically significant, whereas significant differences were observed in nearly all comparisons of Polish and non-Polish European populations. Phylogenetic analyses demonstrated tight clustering of Polish populations separated from non-Polish groups. Population clustering based on Y-STR haplotypes generally correlates well with the geography and history of the region. Thus, our data are consistent with the assumption of homogeneity of present-day paternal lineages within Poland and their distinctiveness from other parts of Europe, at least in respect to their Y-STR haplotypes. Electronic supplementary material to this paper can be obtained by using the Springer LINK server located at <http://dx.doi.org/10.1007/s00439-002-0728-0>.

Introduction

Slavic-speaking populations originate from a territory between the upper Bug river and the mid-Dnepr river. Migration routes of the Slavs led as far as the River Elbe in the West, the River Don in the east, and the Balkans in southeastern Europe. These migrations and the invasions of the Magyars at the end of the ninth century A.D. divided the Slavs into Western, Eastern and Southern populations. The Poles are descendants of Western Slavs who settled the territory between the Rivers Warta and Wisla (Vistula) from the fifth or sixth century (Portal 1969). Historically, the present-day territory of Poland has been

repeatedly settled and resettled by a variety of human populations, notably the Germans (in the twelfth and thirteenth centuries and in the seventeenth to twentieth centuries). A close historical relationship existed with Lithuania, which was unified with Poland from the fourteenth to the eighteenth century, and with Latvia, which was heavily under the influence of the unified Polish-Lithuanian state. Other populations (e.g. Jews, Ukrainians, Belarusians) settled in Poland, which reached its largest extension in the seventeenth century with Eastern borders at the River Dnepr. After that, Poland repeatedly lost large territories to neighbouring countries and, in 1795, disappeared completely from the political map of Europe, not reappearing until after World War I. Later, World War II and its devastating geopolitical consequences in Central and Eastern Europe led among other things to forced displacement of a German population of 8 million in Silesia, Pomerania, and western and eastern Prussia by a population of 3 million Poles, most of whom had formerly been settled in present-day Ukraine and Belarus. More than 2.6 million Polish Jews were exterminated during the German occupation. In 1939 31% of Poland's population was still of non-Polish descent, while today only 450,000 members of populations groups of non-Polish descent are living among nearly 40 million Poles.

To investigate the putatively emerging genetic homogeneity of the present-day Polish population, we have performed a population genetic analysis on the basis of Y-chromosomal microsatellite haplotypes. Microsatellite or short tandem repeat (STR) sequences located on the human Y chromosome have been described as sensitive tools that can be used to characterize even phylogenetically closely related neighbouring populations, such as the Germans and the Dutch (Kayser et al. 2002; Roewer et al. 1996, 2001). They are also commonly used in differentiating male lineages within regional populations, for instance in paternity testing and forensic analysis (Kayser et al. 1997, 2001b; Roewer et al. 2001). The relatively high mutation rates at these rapidly evolving loci (Kayser et al. 1997, 2000a) correlate with an extensive local polymorphism, which allows analyses of migrations and settlements in historical rather than evolutionary time spans (de Knijff et al. 1997; Kayser et al. 2001b). Thus, an analysis of molecular variance (AMOVA) approach based on Y-STR haplotype data is expected to be the method of choice to investigate the degree and significance of differentiation of present-day Polish and other Eastern and central European male populations that shared territories and interacted for quite a long period of historical time.

Materials and methods

Samples

The 919 unrelated male Polish individuals analysed were sampled from six regional populations in Poland: from Warsaw ($n=240$), Bydgoszcz ($n=167$), Gdansk ($n=150$), Lublin ($n=134$), Wrocław ($n=121$), and Krakow ($n=107$). For comparative reasons 1,272 individuals from nine eastern and southern European populations were also included: Germans from Berlin ($n=239$) and from

Leipzig ($n=200$), Russians from Moscow ($n=85$), Lithuanians from Vilnius ($n=152$), Estonians from Tartu ($n=133$), and Latvians from Riga ($n=145$). Samples came from the respective towns and the surrounding areas. In addition, published Y-STR haplotype data of Hungarians living in Budapest ($n=115$) and of Baranya-Romanians ($n=78$) (Füredi et al. 1999), and also of Italians from the area in and around Rome ($n=125$) (Caglia et al. 1998) were included. The geographical locations of the populations studied are shown and the relevant rivers are indicated in Fig. 1.

Genetic analysis

Nine Y-chromosomal microsatellite or short tandem repeat (STR) loci have been analysed: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385. Locus information and PCR-primer sequences can be obtained from Kayser et al. (1997) or from the website: <http://www.ystr.org>. Consistent allele designation and typing quality were assured by simultaneous electrophoretic analysis of sequenced allelic ladders or sequenced reference DNA samples. In addition, all laboratories have successfully passed genotyping quality control tests, e.g. the test evaluated / certified by the Institute of Legal Medicine, Humboldt-University, Berlin (<http://www.ystr.org/europe>). The outline of the laboratory procedures used by different laboratories is given in Table 1 (detailed protocols are available on request).

Statistical Analyses

Haplotype diversity, mean number of pairwise differences, R_{ST} values and associated probability values estimated from 10,000 permutations were calculated, and AMOVA was performed based on the Y-STR haplotypes using the software package ARLEQUIN (<http://anthropologie.unige.ch/arlequin>). A neighbour-joining tree was produced from the pairwise R_{ST} values using the relevant programs in PHYLIP (<http://evolution.genetics.washington.edu/phylip.html>) and viewed using the program TREEVIEW (<http://taxonomy.zoology.gla.ac.uk/rod/rod.html>). A multidimensional scaling analysis based on the pairwise R_{ST} values was performed using the commercially available software package STATISTICA (Statsoft). Differences between diversity values (haplotype diversity and pairwise differences) of the average Polish group and the average non-Polish group were tested for significance by using the single population values and applying a generalized Student's *t*-test in order to account for the observed differences in standard deviations (Welch 1947).

In all statistical analyses, alleles at DYS389II were considered excluding variation at DYS389I. For DYS385, because of the unavailability of a separate genetic analysis of this putatively duplicated Y-STR system, the allele locus assignment was performed so that for each individual the smaller allele was referred to one (DYS385a) and the longer to the other (DYS385b) locus. We are aware of the potential source of uncertainty caused by this procedure. However, the major conclusions of the paper are not influenced by this procedure since (1) identical significance patterns in the diversity comparisons, (2) basically identical significance patterns in pairwise R_{ST} comparisons, and (3) basically the same phylogenetic relationships were revealed when analyses were repeated but without the DYS385 data.

Results

Among the 919 Polish males studied a total of 562 different nine-locus Y-STR haplotypes were observed. The most frequent haplotype occurred 41 times (4.5%), while the four next most frequent haplotypes were found 31, 15, 13 and 10 times (with frequencies of 3.4%, 1.6%, 1.5% and 1.1%, respectively). The remaining haplotypes occurred

Fig. 1 Map of Europe with the location of the populations studied in Poland: Bydgoszcz (1), Krakow (2), Gdansk (3), Wroclaw (4), Warsaw (5), Lublin (6); Russia: Moscow (7); Lithuania: Vilnius (8); Latvia: Riga (9); Estonia: Tartu (10); Germany: Berlin (11) and Leipzig (12); Hungary: Budapest (13) and Hungarian-Romany from Baranya county (14); and Italy: Rome (15). Rivers mentioned in the text are indicated

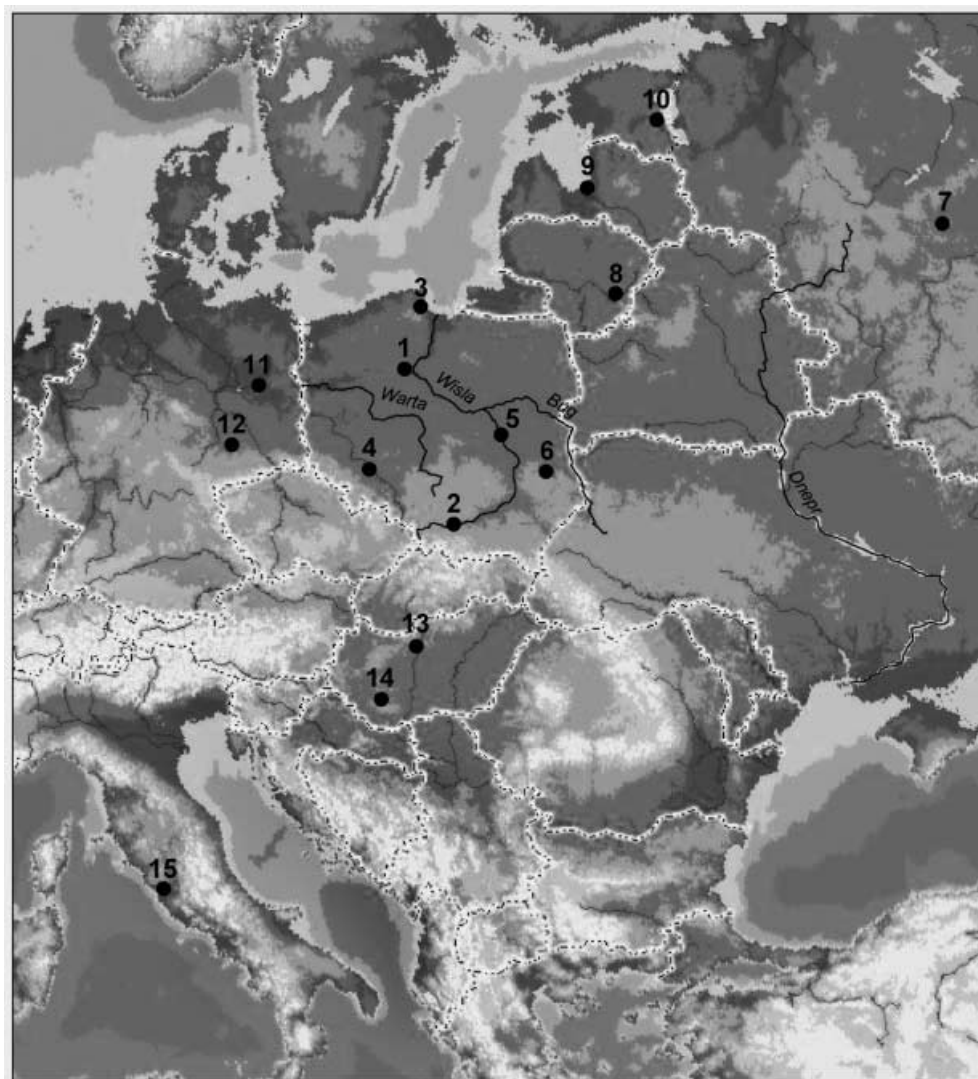


Table 1 An outline of the typing procedures used by different laboratories

Population	PCR amplification	Separation / detection
Bydgoszcz	DYS19, DYS390, DYS389I/II in quadruplex; DYS391, DYS392, DYS393 amplified in triplex; DYS385 in single analyses	ABI 377
Gdańsk	Single analyses	ABI 310
Lublin	DYS19, DYS391, DYS392, DYS393 in quadruplex; DYS385/I and II, DYS389 I/ II, DYS390 in triplex	Separation: 4% denaturing PAGE Detection: FMBIO II scanner (Hitachi)
Germans / Wroclaw	DYS389I/II, DYS390, DYS385 amplified in pentaplex; DYS391, DYS392, DYS393, DYS19 in quadruplex	ABI 377
Russians	(Bosch et al. 2002)	ABI 377
Warsaw	Single analyses	5% denaturing PAGE, silver staining
Krakow	(Kupiec et al. 2000)	ABI 310
Baltic States	(Lessig et al. 2001)	ABI 310

with frequencies of less than 1%. Within the 919 males there were 450 Y-STR haplotypes (80.1%) that were found only once each.

Considerably high Y-STR haplotype diversity (>0.99) was observed in every regional population except the endogamous community of the Hungarian Baranya-Roma-

nies, as was a high mean number of mean pairwise differences (>7) (Table 2). Consideration of the pooled Polish and non-Polish populations revealed smaller diversity values for the Polish sample at both diversity indices (Table 2). A statistical test revealed significant differences ($P < 0.01$) in the mean number of pairwise differences be-

Table 2 Diversity of Y-STR haplotypes based on nine loci in six Polish and nine additional European populations ($N=2,191$)

Population (country)	No. of individuals	No. of haplotypes	Haplotype diversity \pm SD	Mean no. of pairwise differences \pm SD
Bydgoszcz (Poland)	167	135	0.9953 \pm 0.0019	7.66 \pm 3.92
Krakow (Poland)	107	87	0.9931 \pm 0.0033	8.06 \pm 3.81
Gdańsk (Poland)	150	113	0.9933 \pm 0.0024	7.90 \pm 4.19
Wroclaw (Poland)	121	98	0.9910 \pm 0.0038	7.23 \pm 3.93
Warsaw (Poland)	240	180	0.9944 \pm 0.0016	7.86 \pm 3.94
Lublin (Poland)	134	125	0.9985 \pm 0.0013	7.81 \pm 3.58
Moscow (Russia)	85	68	0.9916 \pm 0.0042	8.23 \pm 4.20
Vilnius (Lithuania)	152	123	0.9956 \pm 0.0016	8.14 \pm 3.85
Riga (Latvia)	145	120	0.9960 \pm 0.0017	8.49 \pm 4.11
Tartu (Estonia)	133	106	0.9949 \pm 0.0019	8.72 \pm 4.13
Berlin (Germany)	239	191	0.9966 \pm 0.0010	9.03 \pm 3.96
Leipzig (Germany)	200	164	0.9960 \pm 0.0014	8.46 \pm 4.03
Budapest (Hungary)	115	107	0.9988 \pm 0.0013	9.70 \pm 3.50
Rome (Italy)	125	121	0.9995 \pm 0.0011	10.28 \pm 3.80
Romany (Hungary)	78	32	0.9234 \pm 0.0195	9.25 \pm 5.23
All Polish regions	919	562	0.9950 \pm 0.0007	7.76 \pm 3.90
All non-Polish regions	1,272	835	0.9977 \pm 0.0002	9.46 \pm 4.03

tween the average Polish and the average non-Polish group (both with and without the Romany), whereas non-significant differences were obtained based on the haplotype diversity values (with and without the Romanies). The outlying position of the Baranya-Romanies has been discussed in detail elsewhere (Füredi et al. 1999).

Analysis of molecular variance (AMOVA) revealed that when the Polish dataset was considered exclusively no molecular variation was evident among the different regions of Poland, and thus all of the variance was found within the regional populations (Table 3). This is in contrast to findings in the non-Polish dataset, where a significant proportion of 9.3% ($P<0.0001$) was due to variation among populations and 90.7% was within populations. When Polish and non-Polish populations were grouped for analysis, 4.6% ($P<0.01$) was attributed to variation among those two groups (Table 3).

Consequently, the pairwise population comparisons of R_{ST} revealed that for the six Polish populations values

were close to zero and thus not statistically significant ($P>0.05$), indicating that there was no substructure of male lineages based on Y-STR haplotypes within Poland. In contrast, comparisons of the Polish with non-Polish European populations showed statistically significant differences at the 5% level in all 54 pairwise tests, at the 1% level in 53 of the 54 tests, and at the 0.1% level in 47 of the 54 tests. Five out of the seven tests with non-significant ($0.001>P<0.01$) results between the Polish and the non-Polish groups included the Russian sample, and one each of them included the Lithuanian and the Latvian populations (Table 4). When samples from all six Polish regions were analysed in a pooled fashion and compared with non-Polish Europeans, R_{ST} -values were always statistically significant ($P=0.0004$ for comparison with Russians and $P<0.0001$ for all other pairwise comparisons). Among non-Polish European populations R_{ST} -values were statistically significant at the 5% level in 34 of the 36 comparisons (not between the Lithuanians and Lat-

Table 3 Results from analysis of molecular variance (AMOVA)^a

Dataset	Grouping	Source of variation	Variation (%)	Significance (P -value)
All Polish regions	No	Among populations	0	0.6338
		Within populations	100	
All non-Polish regions ^b	No	Among populations	9.3	<0.0001
		Within populations	90.7	
All regions ^b	No	Among populations	8.4	<0.0001
		Within populations	91.6	
All regions ^b	Polish versus non-Polish	Among groups	4.6	0.0033
		Among populations within groups	5.8	<0.0001
		Within populations	89.6	<0.0001

^aDistance method: sum of squared size difference between Y-STR haplotypes (R_{ST})

^bExclusion of the Romany group did not change the results significantly

Table 4 Pairwise R_{ST} -values (and their P-values^a) below the diagonal and pairwise number of shared Y-STR haplotypes (%) above the diagonal

	Bydgoszcz	Krakow	Gdansk	Wroclaw	Warsaw	Lublin	Moscow	Vilnius	Riga	Tartu	Berlin	Leipzig	Budapest	Rome	Romany
Bydgoszcz	–														
Krakow	-0.0041 (0.7719)	–													
Gdansk	-0.0015 (0.5228)	-0.0024 (0.5750)	–												
Wroclaw	-0.0003 (0.3949)	0.0103 (0.0722)	0.0078 (0.0871)	–											
Warsaw	-0.0038 (0.9474)	-0.0049 (0.9367)	-0.0016 (0.5747)	0.0065 (0.0809)	–										
Lublin	-0.0039 (0.8419)	-0.0038 (0.7204)	-0.0050 (0.9365)	0.0016 (0.2873)	-0.0027 (0.7249)	–									
Moscow	0.0366 (0.0016)	0.0176 (0.0343)	0.0272 (0.0058)	0.0709 (0.0000)	0.0241 (0.0053)	0.0305 (0.0037)	–								
Vilnius	0.0369 (0.0000)	0.0483 (0.0000)	0.0581 (0.0000)	0.0289 (0.0012)	0.0419 (0.0000)	0.0474 (0.0000)	0.0915 (0.0000)	–							
Riga	0.0292 (0.0004)	0.0312 (0.0014)	0.0441 (0.0001)	0.0316 (0.0008)	0.0299 (0.0003)	0.0345 (0.0000)	0.0514 (0.0001)	0.0032 (0.1875)	–						
Tartu	0.1052 (0.0000)	0.0987 (0.0000)	0.1217 (0.0000)	0.1193 (0.0000)	0.0990 (0.0000)	0.1111 (0.0000)	0.0839 (0.0000)	0.0532 (0.0000)	0.0261 (0.0016)	–					
Berlin	0.0528 (0.0000)	0.0408 (0.0000)	0.0557 (0.0000)	0.0753 (0.0000)	0.0444 (0.0000)	0.0546 (0.0000)	0.0173 (0.0142)	0.0605 (0.0000)	0.0272 (0.0002)	0.0264 (0.0006)	–				
Leipzig	0.0708 (0.0000)	0.0570 (0.0000)	0.0749 (0.0000)	0.0970 (0.0000)	0.0610 (0.0000)	0.0742 (0.0000)	0.0327 (0.0023)	0.0734 (0.0000)	0.0362 (0.0002)	0.0314 (0.0004)	-0.0008 (0.4953)	–			
Budapest	0.1157 (0.0000)	0.0848 (0.0000)	0.0983 (0.0000)	0.1579 (0.0000)	0.0982 (0.0000)	0.1053 (0.0000)	0.0197 (0.0176)	0.1692 (0.0000)	0.1153 (0.0000)	0.1144 (0.0000)	0.0357 (0.0003)	0.0442 (0.0000)	–		
Rome	0.2464 (0.0000)	0.2073 (0.0000)	0.2230 (0.0000)	0.2856 (0.0000)	0.2310 (0.0000)	0.2310 (0.0000)	0.1201 (0.0000)	0.2849 (0.0000)	0.2236 (0.0000)	0.1916 (0.0000)	0.1175 (0.0000)	0.1188 (0.0000)	<u>0.0402</u> (0.0005)	–	
Romany	0.2416 (0.0000)	0.1993 (0.0000)	0.2079 (0.0000)	0.2892 (0.0000)	0.2217 (0.0000)	0.2191 (0.0000)	0.1082 (0.0000)	0.3103 (0.0000)	0.2512 (0.0000)	0.2521 (0.0000)	0.1533 (0.0000)	0.1742 (0.0000)	0.0459 (0.0017)	0.0457 (0.0019)	–

^aNon-significant R_{ST} values are *bold and underlined* (5% level), *bold* (1%), *bold and italic* (0.1%), or *underlined* (0.048% = 5% after Bonferroni correction for multiple tests)

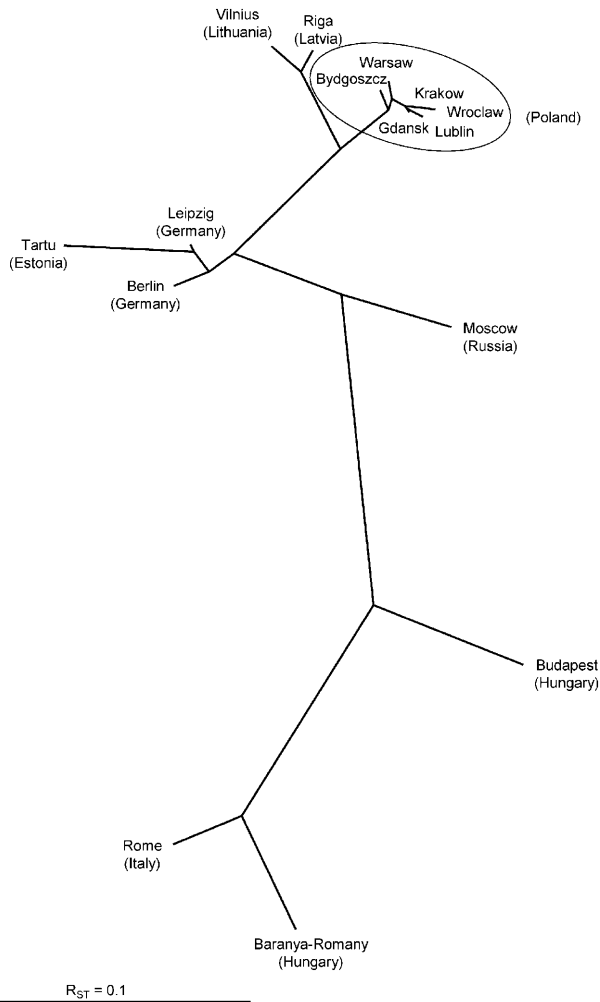
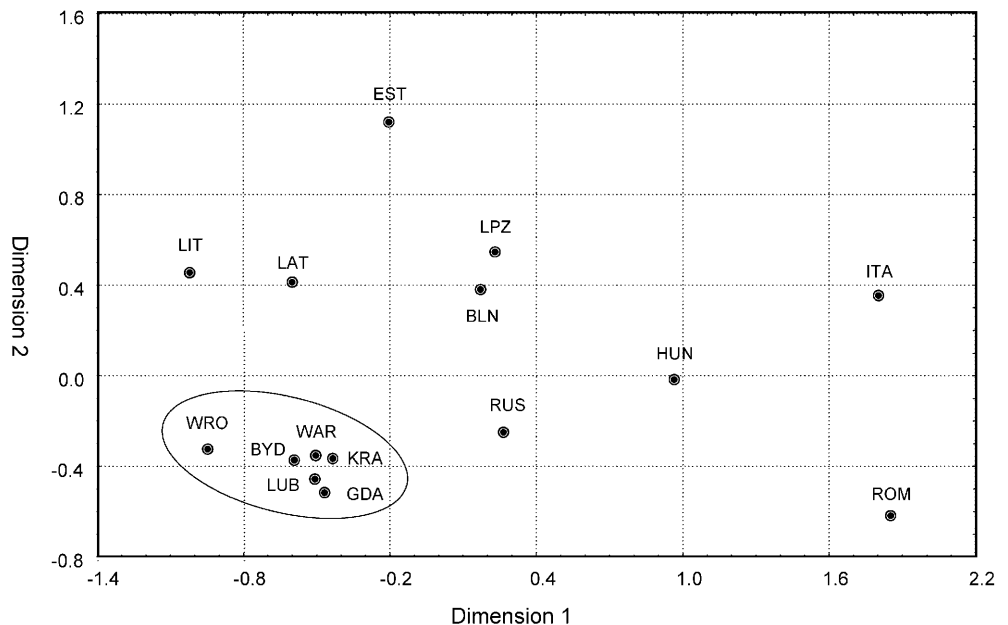


Fig. 2 Neighbour-joining tree based on pairwise R_{ST} values from Y-STR haplotypes of six Polish and nine additional European populations. Polish populations are highlighted

Fig. 3 Two-dimensional plot from multidimensional scaling analysis based on pairwise R_{ST} values from Y-STR haplotypes of six Polish: Bydgoszcz (BYD), Krakow (KRA), Gdansk (GDA), Wroclaw (WRO), Warsaw (WAR), Lublin (LUB), and nine additional European populations: Russians from Moscow (RUS), Lithuanians from Vilnius (LIT), Latvians from Riga (LAT), Estonians from Tartu (EST), Germans from Berlin (BLN) and Leipzig (LPZ), Hungarians from Budapest (HUN), Romany from Hungary (ROM), Italians from Rome (ITA). Polish populations are highlighted



vians or between the two German groups), at the 1% level in 4 of 36 tests, and at the 0.1% level in 28 of 36 tests (Table 4).

The relative distances among the populations studied (as measured by pairwise R_{ST} -values) are displayed graphically in Figs. 2 (as a neighbour-joining tree) and 3 (as a two-dimensional plot derived from multidimensional scaling analysis). In both analyses all six Polish populations are tightly clustered and separate from all non-Polish populations. Lithuanians and Latvians appear closest to the Polish populations; somewhat further away are Estonians, Russians and Germans, whereas Hungarians and Italians and the Baranya-Romany are the furthest distant.

Discussion

We analysed Y-STR haplotypes in six Polish populations from different regions of the country (919 males) and nine populations (1,273 males) from other European countries (in total 2,191 individuals) in order to investigate the amount of genetic heterogeneity and the degree of relatedness among the Poles. Our data are consistent with the assumption of genetic homogeneity of paternal lineages in present-day Poland in respect to Y-STR haplotypes. This is indicated by the AMOVA results, which show no molecular variation among Polish populations, meaning almost zero, and thus statistically non-significant ($P > 0.05$), population differentiation between the different Polish regions according to pairwise R_{ST} , even though the sampled regions are up to 1,000 km apart. The genetic homogeneity of Polish paternal lineages revealed is most probably due to a homogeneous genetic substrate of the ancestral Slavic population, the loss of a considerable amount of both major and minor “ethnic” groups from Poland’s ter-

territory after World War II, and/or the extensive mixing of Poles after World War II due to politically forced resettlement.

The observed homogeneity of paternal lineages within Poland contrasts strikingly with the revealed statistically significant differences between the Polish and the vast majority of non-Polish European populations studied. Evidence for the distinctiveness of Polish paternal lineages compared with other parts of Europe is also evident from a comparison with a large database of European Y-STR haplotypes (European YHRD). The most common Polish haplotype from our study (41/919: 4.5%) occurs in only 21 out of 8,170 (0.29%) non-Polish Europeans from 62 different regions sampled in the European YHRD (<http://www.ystr.org/europe>) as of February 2002. Similarly, the second most common Polish Y-STR haplotype (31/919: 3.4%) is found in 44 of the 8,170 non-Polish Europeans (0.54%). On the other hand, the most common non-Polish European Y-STR haplotype from YHRD (284/8170: 3.5%), which was shared between 50 European populations, was observed in only 6 out of 919 Poles (0.65%), and the second common non-Polish European haplotype (132/8170: 1.6%), shared between 41 European populations, exists only in 4 out of 919 Poles (0.44%).

Our phylogenetic analyses suggest that populations from Latvia and Lithuania were more closely related to the Polish population than any other European groups studied (Figs. 2, 3), although the small R_{ST} -values between these and the Polish groups (0.0289–0.0581) were statistically non-significant in only two pairwise comparisons. These genetic similarities may be the result of admixture due to geographical proximity and/or the tight political links between these countries from the fourteenth to the eighteenth century.

Population samples from Germany and Russia also showed similarities to Polish populations, with relatively small R_{ST} -values on pairwise comparisons (0.0176–0.097). It is noteworthy that all but one of the comparisons between the six Polish populations and the Russians revealed statistically non-significant differences (0.05 < P < 0.001). These genetic similarities are most probably a result of the common Slavic origin. On the other hand, small genetic distances between all of the Polish–German population pairs were statistically significant (P < 0.0001), which might reflect the different background of Slavic-speaking and German-speaking populations. The significant differences revealed between Polish and German samples are especially striking, since the two populations have had close contact during the last millennium and both have inhabited the territory of present-day Poland. This demonstrates a continuous lack of admixture between Germans and Poles, most probably for social, religious and cultural reasons. Genetic difference between Germans and Poles have been reported previously, based on a 1-bp deletion at the Y-chromosomal marker M17 (haplotype Eu19; Semino et al. 2000), which has a high frequency in Poles (56%) but a much lower frequency in Germans (6%). However, other studies, using the Y-SNP marker SRY-1532b (synonym SRY 10831b, haplogroup 3),

which characterises basically the same Y chromosome lineage (Tyler-Smith 1999; Wheale et al. 2001; The Y Chromosome Consortium 2002), have found a much higher frequency of ~30% in larger samples from Germany (M. Kayser, unpublished data; Rosser et al. 2000; Zerjal et al. 1999), which is still only about half the frequency in Poland.

The only two pairs of populations besides the Polish groups that also show non-significant differentiation based on the 5% significance level are the two German populations from Berlin and Leipzig on the one hand and the two Slavic-speaking Baltic populations, the Lithuanians and Latvians on the other (R_{ST} : –0.0008, P > 0.05 and 0.0032, P > 0.05, respectively). The latter observation confirms a previous finding based on independent Lithuanian and Latvian population samples and using five Y-STRs (Zerjal et al. 2001). Interestingly, we find comparatively low population differentiation between the Uralic-speaking Estonians and their Slavic-speaking western neighbours the Latvians (R_{ST} = 0.0261; P = 0.0016; a significant difference of P < 0.001 reported by Lessig et al. (2001) for the same samples is based on fewer Y-STR loci), which is in agreement with previously reported non-significant differences between independent samples from Estonia and Latvia (Zerjal et al. 2001). This might indicate male admixture across linguistic borders, but contrasts with a genetic boundary, which has been identified between Estonians and Latvians based on Y-SNP haplogroup frequencies (Zerjal et al. 2001). However, in the Y-STR-based phylogenetic analyses of both studies, Estonians appear somewhat distant from the Latvians. On the other hand, highly significant differences (P < 0.0001) were revealed between the Estonians and all other geographic neighbours, including the Russians, all Polish groups and also the Lithuanians, a result that is in agreement with linguistic evidence.

Hungarians, Baranya-Romanians and Italians appeared to be most distant from Polish and neighbouring populations in all statistical analyses, reflecting their different population history, geographic locations and linguistic affiliations.

Furthermore, our data indicate that the diversity of Y-STR haplotypes in the Polish population is smaller than in other European groups (except the Romanians). Although the haplotype diversity values were not significantly different from each other, the mean number of pairwise differences of the average Polish group was significantly smaller than that of the average non-Polish group. The mean number of pairwise differences might be seen as a more appropriate measure of the diversity of Y-STR haplotypes, since it takes account of the stepwise mutational process of Y-STRs (Kayser et al. 2000a) by considering the mutational distance between the haplotypes. Haplotype diversity considers only the frequencies of the different haplotypes and thus does not take account of how much they differ from each other. The reduced diversity, together with the observed genetic homogeneity within Poland, could probably be explained by a potentially homogeneous ancestral Slavic population. An alternative ex-

planation postulating a population bottleneck in Polish history might be less plausible, since the largest historically documented population contractions caused by wars within the period 1648–1660 was only 25%. Furthermore, populations who are widely assumed to have gone through a bottleneck, such as the Finns or the Polynesians, show a much more markedly reduced Y-STR diversity than has been observed here for the Polish population, with haplotype diversity values of 0.83 (from five Y-STRs) for the Finns (Zerjal et al. 2001), and 0.81 and a mean number of pairwise differences of 1.55 (from seven Y-STRs) for the Cook Islanders from Polynesia (Kayser et al. 2000b, 2001a).

Homogeneity of paternal lineages based on Y-STR haplotypes as observed here between populations from various Polish regions has also recently been found between a large number of populations in central, western and northern Europe, including 11 different regions of Germany, 5 regions of Holland, 4 of Spain, 6 of Norway, and also populations in Switzerland, Austria, Belgium, and Portugal. All these populations show non-significant ϕ_{ST} values, with $P > 0.05$ in pairwise comparisons (Roewer et al. 2001). Thus, the significant differentiation of Polish, Baltic, and some other eastern European populations observed here and elsewhere (Roewer et al. 2001) clearly demonstrates a somewhat sharp change of paternal lineage composition between the central and the eastern parts of Europe, at least as identified by Y-STR haplotypes. This most probably reflects the different history of those regions and the distinctiveness of Slavic / Baltic and other eastern European male lineages, e.g. those characterized by the Y-SNP haplogroups 3 / Eu19 and 16 (Rosser et al. 2000; Semino et al. 2000), compared with the rather homogeneous central part of Europe.

In conclusion, this study provides the first comprehensive Y chromosome analysis for Poland. Our data indicate that in respect of Y-STR haplotypes the paternal lineages of Poland are genetically rather homogeneous, whereas comparisons with neighbouring populations show similarities and differences that generally correlate well with the geography and history of the region. It would be interesting to include additional eastern European populations, such as Ukrainians, Belarusians, Czechs and Slovaks in a future study. The Y-STR haplotype data used in this study will be included in the European YHRD (<http://www.ystr.org/europe>), allowing frequency estimation of Y-STR haplotypes, and the raw data are available in the appendix.

Acknowledgements We thank Silke Brauer for expert technical assistance, Sandor Füredi for an electronic version of published Y-STR haplotype data, and Arkadiusz Soltysiak and Piotr Jaskulski for helpful comments. General support from Mark Stoneking is gratefully acknowledged, as is statistical advice from Gunter Weiss. M.K. was supported by the Max Planck Society, M.W. by the Ludwik Rydygier Medical University (grant number BW 75/01), and E.B. by the Wellcome Trust. M.A.J. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (grant no. 057559). Finally, we would like to express especial thanks to one of our anonymous reviewers for constructive comments on an earlier version of the manuscript.

References

- Bosch E, Lee AC, Calafell F, Arroyo E, Henneman P, de Knijff P, Jobling MA (2002) High-resolution Y chromosome typing: 19 STRs amplified in three multiplex reactions. *Forensic Sci Int* 125:42–51
- Caglia A, Dobosz M, Boschi I, d'Aloja E, Pascali VL (1998) Increased forensic efficiency of a STR-based Y-specific haplotype by addition of the highly polymorphic DYS385 locus. *Int J Legal Med* 111:142–146
- de Knijff P, Kayser M, Caglia 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, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Roewer L (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 110:134–149
- Füredi S, Woller J, Padar Z, Angyal M (1999) Y-STR haplotyping in two Hungarian populations. *Int J Legal Med* 113:38–42
- Kayser M, Caglia 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, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L (1997) Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 110:125–129
- 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 (2000a) 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, Brauer S, Weiss G, Underhill P, Roewer L, Schiefenhövel W, Stoneking M (2000b) Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 10:1237–1246
- Kayser M, Brauer S, Weiss G, Underhill P, Roewer L, Schiefenhövel W, Stoneking M (2001a) Correction: Melanesian origin of Polynesian Y chromosomes. *Curr Biol* 11(2):I–II
- 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 (2001b) An extensive analysis of Y-chromosomal microsatellite haplotypes in globally dispersed human populations. *Am J Hum Genet* 68:990–1018
- Kayser M, Brauer S, Willuweit S, Schädlich H, Batzer MA, Zawacki J, Prinz M, Roewer L, Stoneking M (2002) Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for U.S. populations. *J Forensic Sci* 47(3):513–519
- Kupiec T, Branicki W, van Mersbergen C, van Hemert G, Kloosterman A (2000) Population genetic study of Y chromosome specific STR-Loci in a population sample from Southern Poland. In: Sensabaugh GF, Lincoln PJ, Olaisen B (eds) *Proceedings of the 18th International ISFH Congress, San Francisco 1999*. (Progress in forensic genetics, vol 8) Elsevier Science, Amsterdam, pp 312–314
- Lessig R, Edelmann J, Krawczak M (2001) Population genetics of Y-chromosomal microsatellites in Baltic males. *Forensic Sci Int* 118:153–157
- Portal R (1969) *The Slavs: a cultural and historical survey of the Slavonic peoples*. Weidenfeld and Nicholson, London
- 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, Caglia A, Carracedo A, Corach D, Dekairelle AF, Dobosz T, Dupuy BM, Furedi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kargel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martinez-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
- Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckmann L, Bertranpetit J, Bosch E, Bradley DG, Brede G, Cooper GC, Corte-Real HBSM, de Knijff P, Decorte R, Dubrova YE, Evgrafov O, Gilissen A, Glisic S, Gölge M, Hill EW, Jeziorowska A, Kalaydjieva L, Kayser M, Kravchenko SA, Lavinha J, Livshits LA, Maria S, McElreavy 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, Previdere 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 by language. *Am J Hum Genet* 67:1526–1543
- Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Maik A, Mika B, Primorac D, Santachiara-Benerecetti AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of palaeolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* 290:1155–1159
- The Y Chromosome Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348
- Tyler-Smith C (1999) Y-chromosomal DNA markers. In: Papiha SS, Deka R, Chakraborty R (eds) *Genomic diversity: applications in human population genetics*, Kluwer Academic / Plenum Publishers, New York, pp 65–73
- Welch BL (1947) The generalization of “Student’s” problem when several different population variances are unequal. *Biometrika* 34:28–35
- Wheale 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
- Zerjal T, Pandya A, Santos FR, Adhikari R, Tarazona E, Kayser M, Evgarof O, Singh L, Thanaraj K, Destro-Bisol G, Thomas MG, Qamar R, Mehdi Q, Rosser ZH, Hurles ME, Jobling MA, Tyler-Smith C (1999) The use of Y-chromosomal DNA variation to investigate population history: recent male spread in Asia and Europe. In: Papiha SS, Deka R, Chakraborty R (eds) *Genomic diversity: applications in human population genetics*, Kluwer Academic / Plenum Publishers, New York, pp 91–102
- Zerjal T, Beckman L, Beckman G, Mikelsaar A-V, Krumina A, Kucinska V, Hurles 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