Hum Genet (2005) 117: 428–443
DOI 10.1007/s00439-005-1333-9

ORIGINAL INVESTIGATION

**Manfred Kayser · Oscar Lao · Katja Anslinger
Christa Augustin · Grazyna Bargel · Jeanett Edelmann
Sahar Elias · Marielle Heinrich · Jürgen Henke
Lotte Henke · Carsten Hohoff · Anett Illing
Anna Jonkisz · Piotr Kuzniar · Arleta Lebioda
Rüdiger Lessig · Slawomir Lewicki
Agnieszka Maciejewska · Dorota Marta Monies
Ryszard Pawłowski · Micaela Poetsch · Dagmar Schmid
Ulrike Schmidt · Peter M. Schneider
Beate Stradmann-Bellinghausen · Reinhard Szibor
Rudolf Wegener · Marcin Wozniak
Magdalena Zoledziewska · Lutz Roewer
Tadeusz Dobosz · Rafal Ploski**

Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis

Received: 24 January 2005 / Accepted: 18 April 2005 / Published online: 16 June 2005

Abstract To test for human population substructure and to investigate human population history we have analysed Y-chromosome diversity using seven microsatellites (Y-STRs) and ten binary markers (Y-SNPs) in samples from eight regionally distributed populations from Poland ($n = 913$) and 11 from Germany ($n = 1,215$). Based on data from both Y-chromosome marker systems, which we found to be highly correlated ($r = 0.96$), and using spatial analysis of the molecular variance (SAMOVA), we revealed statistically significant support for two groups of populations: (1) all Polish populations and (2) all German populations. By means of analysis of the molecular variance (AMOVA) we observed a large and statistically significant proportion of 14% (for Y-SNPs) and 15% (for Y-STRs) of the respective total genetic variation being explained between both countries. The same population differentiation was detected using Monmonier's algorithm, with a resulting genetic border between Poland and Germany that closely resembles the course of the political border between

M. Kayser (✉) · O. Lao
Department of Forensic Molecular Biology,
Medical-Genetic Cluster,
Erasmus University Medical Center Rotterdam,
PO Box 1738, 3000 DR Rotterdam,
The Netherlands
E-mail: m.kayser@erasmusmc.nl
Tel.: +31-10-4638073

K. Anslinger · D. Schmid
Institute of Legal Medicine,
University of Munich, Munich, Germany

C. Augustin · S. Elias
Institute of Legal Medicine,
University of Hamburg, Hamburg, Germany

J. Edelmann · R. Lessig
Institute of Legal Medicine,
University of Leipzig, Leipzig, Germany

M. Heinrich · C. Hohoff
Institute of Legal Medicine,
University of Muenster, Muenster, Germany

J. Henke · L. Henke
Institut fuer Blutgruppenforschung,
Cologne, Germany

A. Illing
Department for Evolutionary Genetics,
Max Planck Institute for Evolutionary Anthropology,
Leipzig, Germany

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

A. Maciejewska · R. Pawłowski
Institute of Forensic Medicine,
Medical University Gdañsk, Gdañsk, Poland

M. Poetsch
Institute of Legal Medicine,
University of Greifswald, Greifswald, Germany

U. Schmidt
Institute of Legal Medicine,
University of Freiburg, Freiburg, Germany

P. M. Schneider · B. Strandmann-Bellinghausen
Institute of Legal Medicine,
University of Mainz, Mainz, Germany

R. Szibor
Institute of Legal Medicine,
University of Magdeburg, Magdeburg, Germany

R. Wegener
Institute of Legal Medicine,
University of Rostock, Rostock, Germany

M. Wozniak
Institute of Molecular and Forensic Genetics,
Collegium Medicum, Nicolaus Copernicus University,
Bydgoszcz, Poland

L. Roewer
Institute of Legal Medicine,
Charite—University Medicine, Berlin, Germany

A. Jonkisz · A. Lebioda · M. Zoledziewska · T. Dobosz
Institute of Forensic Medicine,
Medical University Wroclaw, Wroclaw, Poland

G. Bargel · P. Kuzniar · S. Lewicki · R. Ploski
Human Molecular Genetics Lab,
Departments of Forensic Medicine and Pediatrics,
Medical University Warsaw, Warsaw, Poland

both countries. The observed genetic differentiation was mainly, but not exclusively, due to the frequency distribution of two Y-SNP haplogroups and their associated Y-STR haplotypes: R1a1*, most frequent in Poland, and R1*(xR1a1), most frequent in Germany. We suggest here that the pronounced population differentiation between the two geographically neighbouring countries, Poland and Germany, is the consequence of very recent events in human population history, namely the forced human resettlement of many millions of Germans and Poles during and, especially, shortly after World War II. In addition, our findings have consequences for the forensic application of Y-chromosome markers, strongly supporting the implementation of population substructure into forensic Y chromosome databases, and also for genetic association studies.

Introduction

It is often believed that most neutral human genetic variation observed today has its root far back in time and is a result of ancient rather than recent population movements. This has led to a large number of studies in which genetic analysis of contemporary human populations is used to reconstruct ancient human genetic history (Bowcock et al. 1994; Stoneking and Soodyall 1996; Jin and Su 2000; Jobling and Tyler-Smith 2003; Barbujani and Goldstein 2004; Schurr 2004). However, in principle all migration events, recent or ancient, can leave their traces in the genome and thus can influence genetic diversity as observed at a given point in time, if they involve enough individuals of genetically differentiated populations, and/or result in preferential reproduction. Therefore, neutral genetic diversity as observed today can—in principle—be a mixture of an unknown number of population movements in the ancient but also the recent past. The human Y chromosome, due to its mostly non-recombining inheritance and its small effective population size, has been proven to be a good detector of migration events in human population history (Jobling and Tyler-Smith 2003). Y-chromosome DNA analysis has successfully contributed to a better

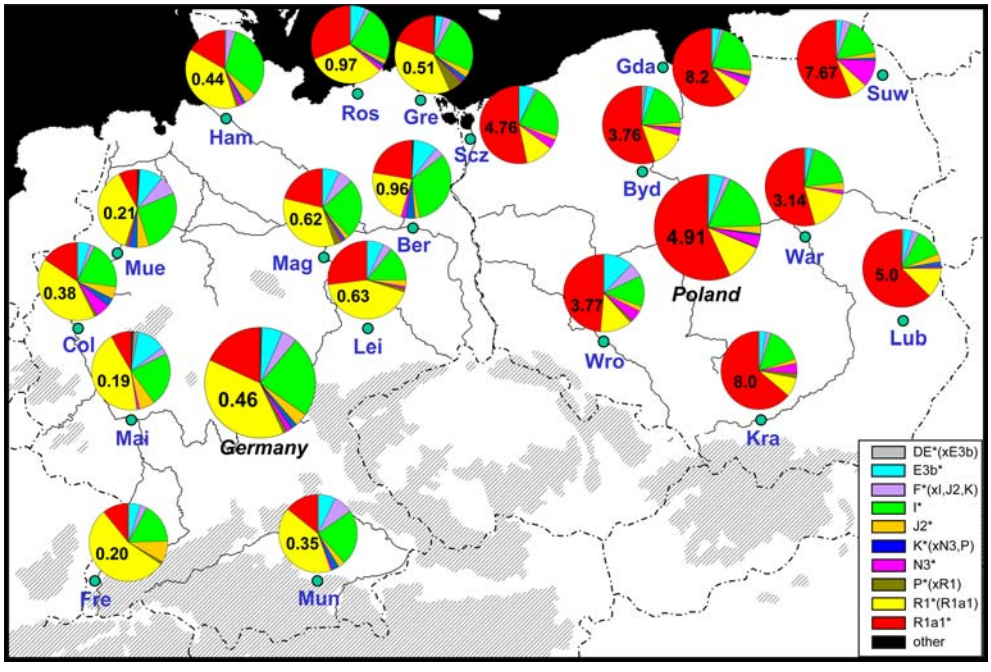

Fig. 1 Haplogroup distribution in regional populations from Poland (eight regions) and Germany (11 regions) and for pooled German and Polish data. *Numbers* indicate the ratio of haplogroup R1a1* to haplogroup R1*(xR1a1). For population abbreviations see Table 2. Note the striking differences in haplogroup R1*(xR1a1) and haplogroup R1a1* distributions (and thus in the ratio) between Polish and German populations

Table 1 The PCR and RFLP typing conditions for ten Y-chromosome binary markers

Marker/mutation	Forward primer (5′ → 3′)	Reverse primer (5′ → 3′)	Annealing temperature (°C)	Enzyme (PCR-RFLP)	PCR-RFLP fragment(s) (bp)	
					Ancestral	Mutant
M9 C → G[a]	GCAGCATATAAAACTTCAGG	GAAATGCATAATGAAGTAAGCG	54	*Hin*fI	100 + 64	164
M74 G → A[b]	AACTAGGAAAGTCTGAAAAATAATCAGA	GCTGCTGTTGTCTTTTAAGTAACTTACT	56	*Rsa*I	151	47 + 104
M170 A → C[c]	TATTTACTTAAAAATCATGGTTC	CCAATTACTTTCAACATTTAAGACC	49	*Nla*IV	99	23 + 76
M173 A → C	TTTCTGAATATTAACAGATGACAACG	CAGTACTCACTTTAGGTTTGCCA	63/56[d]	*Hpy*CH4IV	102 + 26	128
M46 (Tat) T → C[e]	GACTCTGAGTGTAGACTTGTGA	GAAGGTGCCGTAAAAGTGTGAA	60	*Nla*III	85 + 27	112
M172 T → C[f]	TCTCCATCAGAAGATGCCCCAT	ATAATTGAAGACCTTTTAACT	46	*Sml*I	126	104 + 22
M17 G → ins[f]	GTGGTTGCTGGTTGTTACCGG	AGCTGACCACAAACTGATGTAGA	53	*Age*I	124	104 + 19
M35 G → C	TAAGCCTAAAGAGCAGTCAGAG	AGAGGGAGCAATGAGGACA	63/56[d]	*Bts*I	513	351 + 162
M89 C → T[g]	ACAGAAGGATGCTGCTCAGCTT	GCAACTCAGGCAAAGTGAGACAT	56	*Nla*III	65+22	87
YAP del → ins[h]	CAGGGGAAGATAAAGAAATA	ACTGCTAAAAGGGGATGGAT	50	–	150	455

[a]Kayser et al. (2000b)
[b]Kayser et al. (2003)
[c]Nasidze et al. (2004)
[d]Touch-down PCR, decreasing annealing temperature by 0.5°C for 14 cycles, followed by 25 cycles at constant temperature
[e]Zerjal et al. (1997)
[f]Cordaux et al. (2004)
[g]Ke et al. (2001b)
[h]Hammer and Horai (1995)

understanding of the more ancient human population history [i.e. from many thousands of years ago (Rosser et al. 2000; Kayser et al. 2001; Wells et al. 2001; Semino et al. 2002; Zegura et al. 2004)] and the more recent history of human populations [i.e. from a few thousand or some hundreds of years ago (Kayser et al. 2000a, Wilson et al. 2001; Weale et al. 2002; Zerjal et al. 2003; Capelli et al. 2003; Bosch et al. 2003)]. However, studies that convincingly demonstrate the influence of very recent events in human population history (i.e. a few hundred years) to human genetic diversity are rare (Soodyall et al. 2003; Hurles et al. 2004) and do not yet exist for events less than a hundred years.

The Polish population is interesting for studying the effect of population history on human genetic diversity, since it has suffered from a large number of severe changes in its territory in the very recent past, the more distant past, and also in the historical and ancient past, leading to human population movements. In a previous study, we showed that haplotypes defined by Y-chromosome microsatellites [or short tandem repeats (STRs)] were surprisingly homogeneous within Poland, but differed significantly from populations of neighbouring geographic regions (Ploski et al. 2002). In particular, we observed statistically significant Y-STR differences between all six Polish and two German populations studied. Such pronounced differences were unexpected, given the close interactions between the Poles and Germans, such as those caused by the intense German settlements in Silesia and Pomerania in the thirteenth to fifteenth centuries, and the political and social events associated with progressive losses of western Polish territories to the Prussian kingdom in the eighteenth century.

The unique inheritance of the Y chromosome offers the possibility of choosing genetic markers relative to the time scales of the population history event under question, because of their highly different mutation rates. The Y-STRs are believed to be suitable for more recent events, whereas Y-chromosome single-nucleotide polymorphisms (Y-SNPs) are suitable for more ancient events (de Knijff 2000). This has been concluded because of a 100,000-times lower mutation rate of Y-SNPs compared with Y-STRs (Kayser et al. 2000b; Thomson et al. 2000). However, systematic studies to compare the power of both marker systems in detecting the time-depth of human population history by analysing both marker systems in parallel are still scarce.

The purpose of the present study was to investigate in detail the Polish–German differences in male lineages by (1) expanding the population sample and including a systematic representation of Polish as well as German geographic sub-regions, and by (2) analysing Y-chromosomal SNPs—in parallel with Y-STRs—to investigate the time-depth of the Polish–German Y-chromosome differentiation and to evaluate the correlation of the regional differentiation as observed so far based on Y-STRs (Ploski et al. 2002) with the pan-European frequency gradients as reported based on Y-SNPs (Rosser et al. 2000; Semino et al. 2000).

Table 2 Y-SNP haplogroup counts and frequencies (%) in populations studied (and their diagnostic binary markers)

Population	*n*	DE*(xE3b) (YAP)	%	E3b* (M35)	%	F*(xI,J2,K) (M89)	%	I* (M170)	%	J2* (M172)	%	K*(xN3,P) (M9)	%	N3* (M46)	%	P*(xR1) (M74)	%	R1*(xR1a1) (M173)	%	R1a1* (M17)	%	Other (ancestral)	%
Poland																							
Wroclaw (Wro)	101	0	**0**	12	**11.9**	6	**5.9**	13	**12.9**	2	**2.0**	0	**0**	5	**5.0**	1	**1**	13	**12.9**	49	**48.5**	0	**0**
Warsaw (War)	121	1	**0.8**	3	**2.5**	1	**0.8**	23	**19.0**	4	**3.3**	0	**0**	2	**1.7**	0	**0**	21	**17.4**	66	**54.5**	0	**0**
Lublin (Lub)	112	1	**0.9**	4	**3.6**	3	**2.7**	13	**11.6**	4	**3.6**	2	**1.8**	1	**0.9**	0	**0**	14	**12.5**	70	**62.5**	0	**0**
Gdansk (Gda)	150	0	**0**	5	**3.3**	2	**1.3**	32	**21.3**	4	**2.7**	1	**0.7**	5	**3.3**	0	**0**	11	**7.3**	90	**60.0**	0	**0**
Krakow (Kra)	100	0	**0**	3	**3.0**	2	**2.0**	15	**15.0**	2	**2.0**	0	**0**	4	**4.0**	2	**2.0**	8	**8.0**	64	**64.0**	0	**0**
Szczecin (Szc)	105	0	**0**	7	**6.7**	1	**1.0**	23	**21.9**	2	**1.9**	0	**0**	4	**3.8**	0	**0**	12	**11.4**	56	**53.3**	0	**0**
Suwalki (Suw)	82	0	**0**	2	**2.4**	3	**3.7**	13	**15.9**	2	**2.4**	1	**1.2**	9	**11.0**	0	**0**	6	**7.3**	46	**56.1**	0	**0**
Bydgoszcz (Byd)	142	3	**2.1**	5	**3.5**	0	**0**	26	**18.3**	3	**2.1**	1	**0.7**	4	**2.8**	0	**0**	21	**14.8**	79	**55.6**	0	**0**
Poland all	913	5	**0.5**	41	**4.5**	18	**2.0**	158	**17.3**	23	**2.5**	5	**0.5**	34	**3.7**	3	**0.3**	106	**11.6**	520	**57.0**	0	**0**
Germany																							
Berlin (Ber)	103	0	**0**	10	**9.7**	4	**3.9**	33	**32.0**	2	**1.9**	4	**3.9**	2	**1.9**	0	**0**	24	**23.3**	23	**22.3**	1	**1.0**
Leipzig (Lei)	144	0	**0**	10	**6.9**	5	**3.5**	21	**14.6**	4	**2.8**	0	**0**	1	**0.7**	2	**1.4**	62	**43.1**	39	**27.1**	0	**0**
Magdeburg (Mag)	100	0	**0**	7	**7.0**	6	**6.0**	25	**25.0**	2	**2.0**	1	**1.0**	1	**1.0**	3	**3.0**	34	**34.0**	21	**21.0**	0	**0**
Rostock (Ros)	96	0	**0**	6	**6.3**	2	**2.1**	22	**22.9**	2	**2.1**	1	**1.0**	2	**2.1**	0	**0**	31	**32.3**	30	**31.3**	0	**0**
Greifswald (Gre)	104	0	**0**	3	**2.9**	4	**3.8**	25	**24.0**	3	**2.9**	2	**1.9**	1	**1.0**	6	**5.8**	39	**37.5**	20	**19.2**	1	**1.0**
Hamburg (Ham)	161	1	**0.6**	0	**0**	6	**3.7**	51	**31.7**	8	**5.0**	2	**1.2**	3	**1.9**	2	**1.2**	61	**37.9**	27	**16.8**	0	**0**
Muenster (Mue)	102	0	**0**	10	**9.8**	8	**7.8**	27	**26.5**	5	**4.9**	3	**2.9**	1	**1.0**	1	**1.0**	38	**37.3**	8	**7.8**	1	**1.0**
Freiburg i.Br. (Fre)	102	0	**0**	5	**4.9**	3	**2.9**	17	**16.7**	9	**8.8**	0	**0**	0	**0**	1	**1.0**	56	**54.9**	11	**10.8**	0	**0**
Cologne (Col)	96	0	**0**	5	**5.2**	2	**2.1**	19	**19.8**	5	**5.2**	3	**3.1**	6	**6.3**	1	**1.0**	40	**41.7**	15	**15.6**	0	**0**
Mainz (Mai)	95	2	**2.1**	11	**11.6**	3	**3.2**	21	**22.1**	6	**6.3**	0	**0**	1	**1.1**	0	**0**	42	**44.2**	8	**8.4**	1	**1.1**
Munich (Mun)	112	0	**0**	8	**7.1**	9	**8.0**	26	**23.2**	3	**2.7**	3	**2.7**	1	**0.9**	0	**0**	46	**41.1**	16	**14.3**	0	**0**
Germany all	1215	3	**0.2**	75	**6.2**	52	**4.3**	287	**23.6**	49	**4.0**	19	**1.6**	19	**1.6**	16	**1.3**	473	**38.9**	218	**17.9**	4	**0.3**

Table 3 Y-chromosome Y-SNP and Y-STR diversity in populations studied

Region/ population	*n*	No. of haplogroups	Haplogroup diversity	No. of haplotypes	Haplotype diversity	MPD haplotypes
Poland						
Wroclaw	101	8	0.7180 ± 0.0386	79	0.9923 ± 0.0033	5.440 ± 2.447
Warsaw	121	8	0.6394 ± 0.0369	82	0.9886 ± 0.0034	5.513 ± 2.687
Lublin	112	9	0.5817 ± 0.0496	70	0.9786 ± 0.0061	5.029 ± 2.675
Gdansk	150	8	0.5899 ± 0.0384	91	0.9834 ± 0.0045	4.940 ± 2.587
Krakow	100	8	0.5634 ± 0.0533	69	0.9842 ± 0.0052	4.819 ± 2.714
Szczecin	105	7	0.6544 ± 0.0395	72	0.9881 ± 0.0040	5.488 ± 2.636
Suwalki	82	8	0.6480 ± 0.0511	58	0.9877 ± 0.0047	5.490 ± 2.692
Bydgoszcz	142	8	0.6366 ± 0.0363	93	0.9886 ± 0.0031	5.231 ± 2.513
Poland all	913	10	0.6284 ± 0.0153	330	0.9865 ± 0.0012	5.233 ± 2.628
Germany						
Berlin	103	9	0.7875 ± 0.0197	78	0.9899 ± 0.0037	5.979 ± 2.342
Leipzig	144	8	0.7179 ± 0.0244	99	0.9923 ± 0.0021	5.686 ± 2.341
Magdeburg	100	9	0.7756 ± 0.0216	70	0.9875 ± 0.0043	6.014 ± 2.903
Rostock	96	8	0.7480 ± 0.0203	81	0.9932 ± 0.0034	5.990 ± 2.339
Greifswald	104	10	0.7649 ± 0.0248	84	0.9950 ± 0.0023	6.314 ± 2.614
Hamburg	161	9	0.7280 ± 0.0194	120	0.9940 ± 0.0018	5.747 ± 2.335
Muenster	102	10	0.7732 ± 0.0267	66	0.9699 ± 0.0106	5.820 ± 2.491
Freiburg	102	7	0.6544 ± 0.0438	72	0.9854 ± 0.0052	5.347 ± 2.497
Cologne	96	9	0.7599 ± 0.0314	64	0.9767 ± 0.0074	5.460 ± 2.612
Mainz	95	9	0.7373 ± 0.0339	68	0.9886 ± 0.0039	5.576 ± 2.375
Munich	112	8	0.7506 ± 0.0272	83	0.9887 ± 0.0040	5.878 ± 2.426
Germany all	1,215	11	0.7531 ± 0.0075	520	0.9894 ± 0.0010	5.836 ± 2.473

Materials and methods

DNA samples

The DNA samples of an overall 2,128 unrelated male individuals were included in this study, comprising 913 samples from eight different regions in Poland, and 1,215 samples from 11 different regions in Germany (see Fig. 1 for geographic location and Table 2 for sample size per group).

Table 4 The AMOVA results with statistically significant groupings

Source of variation	Percentage of variation	
	Y-SNPs (F_{ST} [a])	Y-STRs (R_{ST} [a])
Poland versus Germany		
Among groups	14.09	15.07
Among populations within groups	0.87	0.55
Within populations	85.05	84.39
Poland		
Among populations	0.32	0.08
Within populations	99.68	99.92
Germany		
Among populations	1.42	1.00
Within populations	98.58	99.00
East versus West Germany		
Among groups	1.04	1.31
Among populations within groups	0.84	0.29
Within populations	98.12	98.41

[a]Distance method applied

Genotyping

Ten Y-chromosomal binary markers, consisting of eight SNPs [M9, M74, M173, M170, M172, M35, M89 (Underhill et al. 2000) and Tat-M46 (Zerjal et al. 1997; Underhill et al. 2000)], one 1-bp deletion [M17 (Underhill et al. 2000)] and one *Alu* insertion/deletion polymorphism [YAP (Hammer 1994)], were selected to be most informative in the European population based on two previous large population studies (Rosser et al. 2000; Semino et al. 2000). YAP (DYS287) was analysed as described elsewhere (Hammer and Horai 1995). For the other markers, simple PCR-RFLP methods were used in order to assure simple analyses (Table 1). Regional samples were mostly typed in regional laboratories, except for Y-SNP analysis of Berlin, typed in M. Kayser's laboratory, Rostock and Cologne typed in R. Ploski's laboratory, and Krakow, Suwalki and Szczecin typed in R. Ploski's lab for Y-STRs and in T. Dobosz's lab for Y-SNPs. Standard PCR conditions were applied in all laboratories as follows, with additional details provided in Table 1: 0.4 μM of each primer, 1× GeneAmp PCR buffer II (Applied Biosystems, Foster City, Calif., USA), 1.5 μM $MgCl_2$, 1 U AmpliTaq Gold DNA polymerase or AmpliTaq DNA polymerase (Applied Biosystems), 0.2 μM dNTPs (Amersham Pharmacia Biotech, Chalfont, UK), 147 μM bovine serum albumin (Sigma, St. Louis, Mo., USA), 10–100 ng DNA and a hot-start PCR of 4 min 95°C initial denaturation (11 min for AmpliTaq Gold DNA polymerase), followed by 30–35 cycles of 30 s at 94°C, 30 s at the locus-specific annealing temperature, and 45 s at 72°C, followed by a final step of 10 min at 72°C. The PCR products were digested using suitable restriction endonucleases (see Table 1) according to the

Table 5 F-statistics from AMOVA with statistically significant groupings

	Y-SNPs (F_{ST} [a])	Y-STRs (R_{ST} [a])
Poland versus Germany		
F_{SC}	0.01009 ($P < 0.00001$)	0.00643 ($P < 0.00001$)
F_{ST}	0.14952 ($P < 0.00001$)	0.15615 ($P < 0.00001$)
F_{CT}	0.14085 ($P < 0.00001$)	0.15069 ($P < 0.00001$)
Poland		
F_{ST}	0.00323 ($P = 0.11632$)	0.00081 ($P = 0.31769$)
Germany		
F_{ST}	0.01416 ($P < 0.00001$)	0.01004 ($P < 0.00001$)
East versus West Germany		
F_{SC}	0.00851 ($P < 0.00001$)	0.00289 ($P < 0.00001$)
F_{ST}	0.01879 ($P = 0.00098$)	0.01590 ($P = 0.08993$)
F_{CT}	0.01037 ($P = 0.00489$)	0.01305 ($P = 0.00196$)

[a]Distance method applied

recommendations of the suppliers. Digested PCR products were visualised in a 3% NuSieve/1% Seakam-agarose gel using ethidium bromide. For some markers (M17, M170, M172, and M173) no restriction enzyme was commercially available for detection and therefore primer induced RFLP assay (PIRA)-PCR assays were designed using the software described by Ke et al. (2001a) (http://cedar.genetics.soton.ac.uk/public_html/primer2.html). In PIRA-PCR, a mismatch is introduced in the 3′ site of the PCR primer placed immediately next to the SNP, resulting in the creation of a restriction site in combination with the SNP sequence. Binary markers were analysed hierarchical according to the Y-chromosome marker phylogeny (Jobling and Tyler-Smith 2003). Some laboratories (Leipzig, Mainz, Warsaw, and Wroclaw) additionally used alternative protocols, as described elsewhere (Bender et al. 2003; Lessig et al. 2005). Data for binary markers are described here for the first time for all samples except for three markers (M46-Tat, M17, and M9) in the samples from Mainz (Bender et al. 2003). Seven Y-chromosomal microsatellites [or short tandem repeats (Y-STRs)], DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, were analysed as mentioned previously (Ploski et al. 2002). The Y-STR data are described here for the first time for all samples except for Warzaw, Leipzig, 123 out of 150 males from Gdansk, 13 out of 142 males from Bydgoszcz (Ploski et al. 2002; Roewer et al. 2005), as well as Berlin, Magdeburg, Rostock, Greifswald, Freiburg, Mainz, Munich, and 37 out of 102 males from Muenster (Roewer et al. 2005).

Statistical analysis

The degree of genetic differentiation between populations was quantified by means of the analysis of molecular variance (AMOVA) using the Arlequin 2000 package (Schneider et al. 2000). This method allows us to define the percentage of the genetic variation that is explained (1) among groups of population defined a priori, (2) between the populations of the same group,

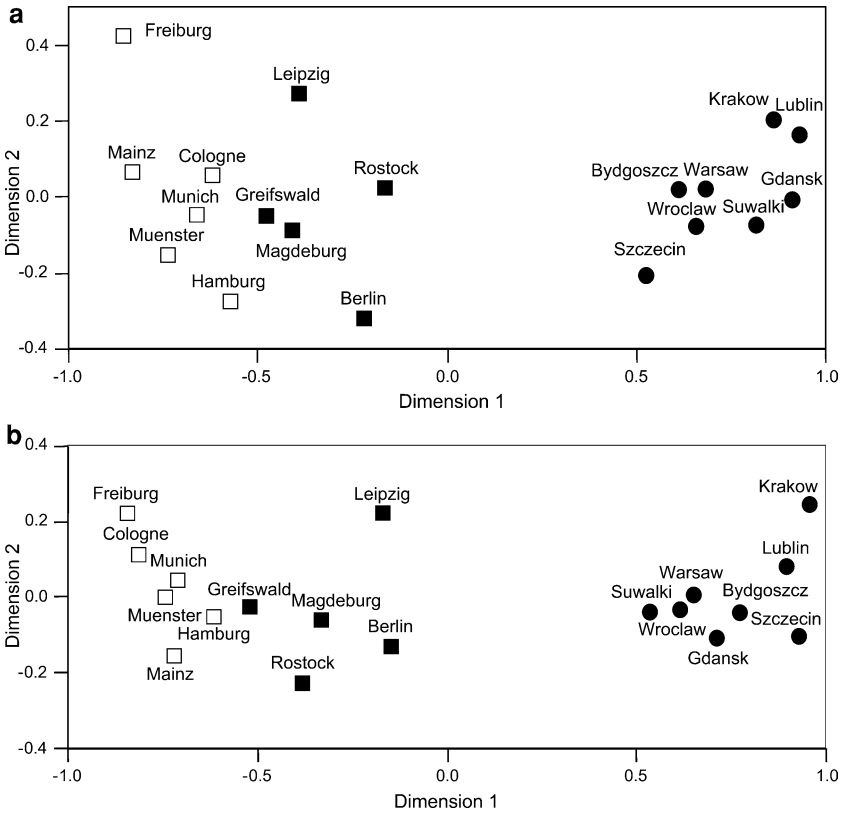

Fig. 2 Plots from multidimensional scaling (*MDS*) analysis of **a** F_{ST} values from Y-SNP haplogroups, stress value 0.03; **b** R_{ST} values from Y-STR haplotypes, stress value 0.03. *Filled circles* indicate Polish populations; *filled squares* indicate Eastern German populations and *empty squares* indicate Western German populations. Note the clear differentiation between Polish and German populations based on both marker systems, and the position of Eastern German populations somewhat between Western German and Polish populations but clearly separated from the latter

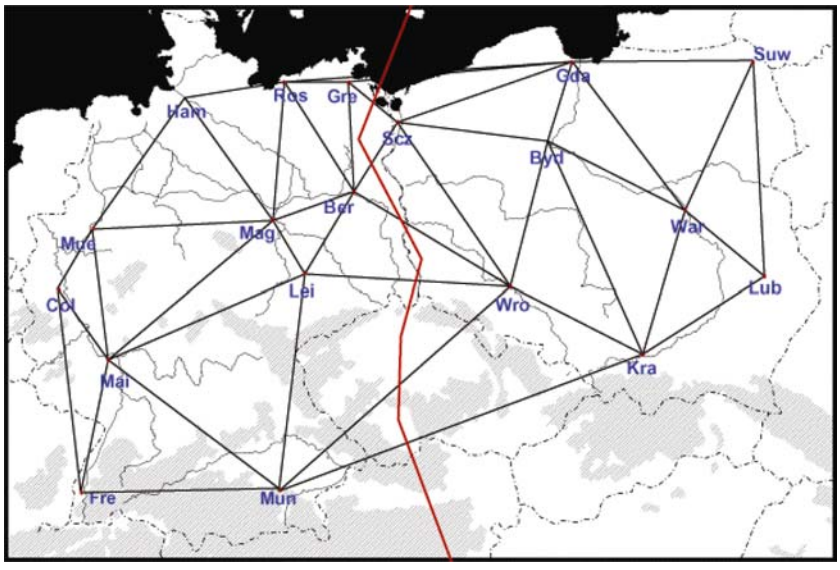

Fig. 3 Barrier analysis based on F_{ST} from Y-SNP haplogroups given the spatial distribution of the populations superimposed on a geographic map (results based on R_{ST} from Y-STR haplotypes are identical). *Red line* indicates the identified genetic barrier. For population abbreviations see Table 2. Note the close resemblance of the course of the genetic border between Polish and German populations with the course of the political border between the two states Germany and Poland

and (3) within the populations. The spatial analysis of the molecular variance (SAMOVA) algorithm (Dupanloup et al. 2002) was used to identify groups of geographically neighbouring populations in order to maximise the genetic differentiation between the groups and minimise the genetic differentiation between the populations within each group; thus, this method allow us to detect the presence of putative geographic barriers between groups of populations based on their genetic diversity. Multidimensional scaling (MDS) analysis was used to plot the pairwise genetic distances F_{ST} based on Y-SNP haplogroups, and R_{ST} based on Y-STR haplotypes (that were computed by means of Arlequin) using the software package SPSS, version 11. This multivariate method defines for each population coordinates so that the distances among them are as close as possible to the original genetic distances. The stress is a measure of goodness-of-fit that indicates how similar is the distance matrix based on the new coordinates to the original genetic distance matrix and it is actually smaller for better fits. Since SPSS converts negative values into missing values, the genetic distances were scaled up to be all-positive. A correspondence analysis was performed with the frequencies of the Y-SNP haplogroups by means of the STATISTICA package. This multivariate method plots in the same graphical representation both columns and rows of a contingence table (in our case, populations and haplogroups based on YSNPs). Plotting both populations and haplogroups in the same graphical representation opens the possibility to asses,

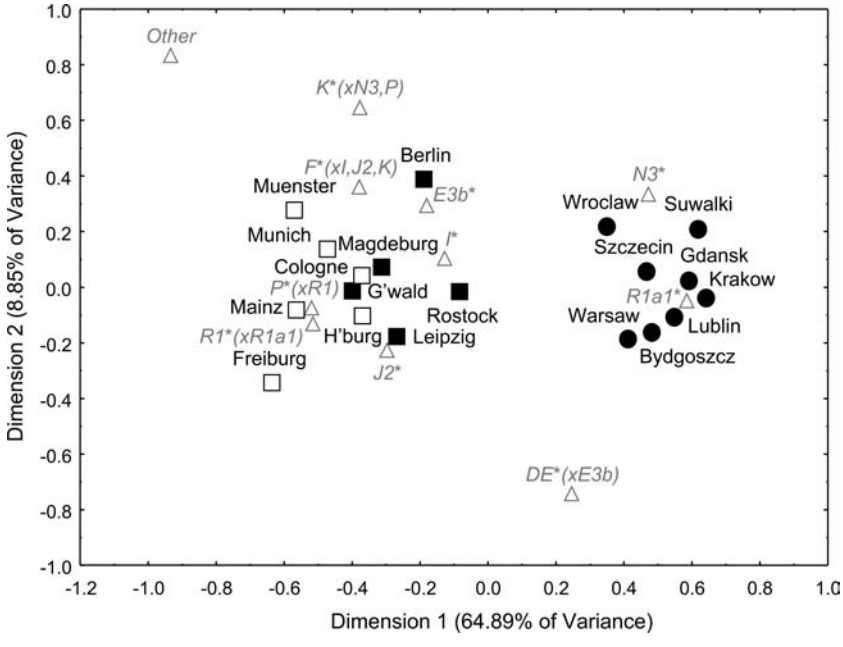

Fig. 4 Correspondence analysis. Two-dimensional plot of the distribution of populations according to their Y-SNP haplogroup frequencies in correspondence with a plot of the haplogroups in the same graphical representation graphical representation. Population designations as in Fig. 2. Positions of haplogroups designated by *triangles* and in *grey*. Note the correspondence between the Polish population cluster and haplogroup R1a1* and N3*, as well as between the German population cluster and haplogroups R1a*(xR1a1), and P*(xR1)

which haplogroups are contributing to the distribution and differentiation of the populations in the plot. The spatial distribution of Y-SNP haplogroups were analysed by means of spatial autocorrelation analysis (Sokal and Oden 1978) using the PASSAGE program (Rosenberg 2001). The spatial autocorrelation analysis computes the level of autocorrelation between pairs of points that are within a certain geographic distance. The plot of the level of autocorrelation in relation to increasing geographic distance classes gives information about the spatial pattern of the data. In the case of a clinal pattern of the data, it is expected that the shape of the autocorrelogram will decrease from positive autocorrelation values for the closest geographical distances to negative values for the longest geographic distance classes (Barbujani 2000). The geographical location of putative genetic barriers was analysed by means of the Barrier version 2.2 program (Manni and Heyer 2004). This program computes Monmonier's algorithm to detect a spatial abrupt rate of change in terms of the genetic differentiation between geographically neighboring populations. Genetic diversity measures [haplotype and haplogroup diversity, number of haplotypes and haplogroups, mean number of pairwise haplotype differences (MPD)] were calculated using the software package Arlequin 2.000 (Schneider et al. 2000). Finally, a general limitation of all relevant analysis dealing with patterns of genetic marker frequencies should be noted here: frequencies of different genetic markers are not independent from each other in the way that a high frequency of one marker in a population consequently leads to a lower frequency of one (or more) different marker(s) in that same population.

Results

Y chromosome diversity

By analysing ten binary markers selected to be most informative in European populations (Semino et al. 2000; Rosser et al. 2000), we were able to identify ten Y-chromosomal haplogroups in 2,128 men from eight different geographic regions in Poland and 11 in Germany (Table 2). Only four out of 2,128 individuals (0.18%, all from Germany) could not be assigned to one of the ten haplogroups (Table 2). All ten Y-chromosome haplogroups were observed in both areas, although the

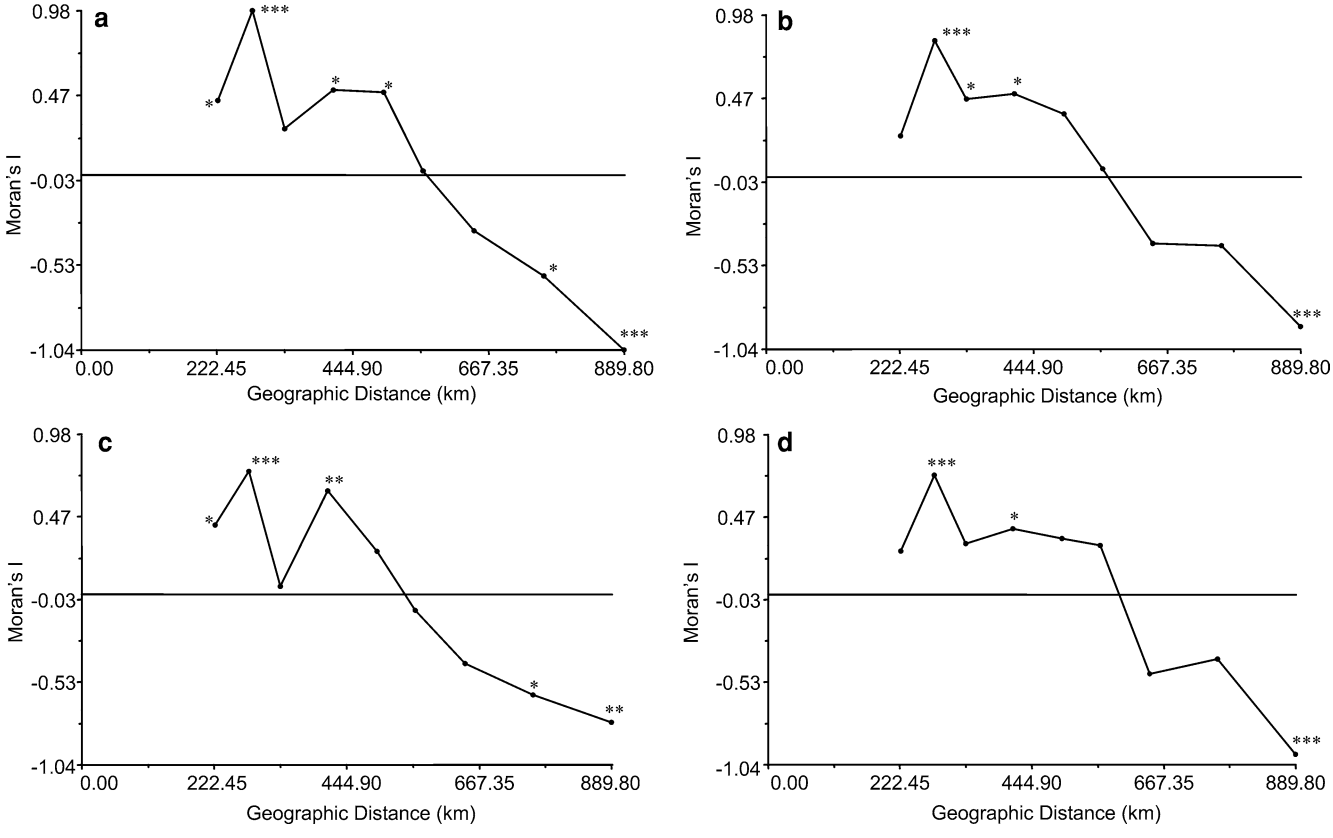

Fig. 5 Graphical representation of spatial autocorrelation analyses for **a** haplogroup R1*(xR1a1); **b** haplogroup R1a1*; **c** haplogroup I1a* using individuals carrying the two most common Y-STR haplotypes inferred to be associated with I1a*; **d** haplogroup I1b* using individuals carrying the two most common Y-STR haplotypes inferred to be associated with I1b*. The *x-axis* represents geographic distance between population samples; the *y-axis* represents Morans's index; *asterisks* indicate the significance of Moran's index, with a *single asterisk* denoting $P < 0.05$, *double asterisks* denoting $0.05 > P > 0.01$, and *triple asterisks* denoting $P < 0.01$. Note the statistically significant results in the long geographic distances but no statistically significant results in short geographic distances

number of haplogroups identified and haplogroup diversity differed between regional populations (Table 3). The total Y-SNP based haplogroup diversity was 0.7563, with a higher diversity in Germany (range: 0.6544–0.7875, pooled: 0.7531) than in Poland (range: 0.5634–0.7180, pooled: 0.6284); the difference between Poland and Germany was statistically significant (Mann–Whitney *U*-test: $Z = -3.428$, $P = 0.001$). The analysis of seven Y-chromosomal microsatellites revealed 705 different haplotypes out of 2,128 individuals (total diversity: 0.9914). The Y-STR-based haplotype diversity was on average almost identical between Germany and Poland (Germany: range: 0.9767–0.9950, pooled: 0.9894 and Poland: range 0.9786–0.9923, pooled 0.9865), and no statistically significant difference was observed (Mann–Whitney *U*-test: $Z = -1.117$, $P = 0.264$). However, the mean number of pairwise differences (MPD) between Y-STR haplotypes was higher in Germany (range: 5.347–6.314, pooled: 5.836) than in Poland (range: 4.819–5.490, pooled: 5.233), and the difference between Poland and Germany was statistically significant (Mann–Whitney *U*-test: $Z = -3.055$, $P = 0.002$). Thus, we observed more Y-chromosome diversity in Germany than in Poland, based on the Y-SNPs and Y-STRs analysed here.

Y-SNP haplogroup distribution

Haplogroup R1*(xR1a1) appeared at the highest frequency in German populations, especially those from Western Germany, whereas haplogroup R1a1* was most frequent in Polish populations (Table 2, Fig. 1). Haplogroup R1*(xR1a1) was on average 3.4-times more frequent in Germany than in Poland, whereas R1a1* was on average 3.2-times more frequent in Poland than in Germany. Differences in R1*(xR1a1) and R1a1* frequencies between German and Polish groups were statistically significant (Mann–Whitney *U*-test: $Z = -3.633$, $P < 0.001$ for both haplogroups). The converse frequency distribution of both haplogroups can be demonstrated by the ratio of haplogroup R1a1* to R1*(xR1a1) (Fig. 1), which was on average more than ten-times higher in Poland (4.91) than in Germany (0.46) and on average twice as high in Eastern (0.65) as in Western Germany (0.30).

Haplogroup I*, the overall third-most frequent haplogroup observed here, was on average 1.4-times more frequent in Germany than in Poland (Table 2, Fig. 1). Differences in hgI* frequencies between German and Polish groups were statistically significant (Mann–Whitney *U*-test: $Z = -2.642$, $P = 0.008$). Although being rare, haplogroup N3* was on average 2.3-times more frequent in Poland than it was in Germany (Table 2, Fig. 1) and the differences between both regions were statistically significant (Mann–Whitney *U*-test: $Z = -2.189$, $P = 0.029$).

The haplogroups J2* and E3b* were on average about 1.5-times more frequent in Germany than in Poland (Table 2, Fig. 1) and haplogroup P*(xR1) was on average 4.3-times more frequent in Germany than in Poland (Table 2, Fig. 1); however, all those haplogroups were overall rather rare and the frequency differences between German and Polish groups were not statistically significant ($P > 0.05$). The remaining haplogroups, DE*(xE3b), F*(xI,J2,K), and K*(xN3,P), are not necessarily representing monophyletic groups—given the selection of markers analysed here—and potentially contain a number of different haplogroups. Differences between Polish and German groups were statistically significant for F*(xI,J2,K) ($Z = -2.396$, $P = 0.017$), but not for DE*(xE3b), and K*(xN3,P) ($P > 0.05$).

Genetic differentiation

In order to test for geographical population substructure in our overall Polish/German Y-chromosome dataset, we performed SAMOVA separately for the Y-SNP and Y-STR data. Based on SAMOVA, two groups of populations were significantly supported by both datasets: on one hand, all German populations and on the other hand all Polish populations. We observed a high and statistically significant level of 14% of the total genetic variation being explained between the Polish and the German group of populations ($P < 0.00001$) based on Y-SNP haplogroups and similarly 15% ($P < 0.00001$) based on Y-STR haplotypes (Tables 4, 5). This clearly demonstrates a strong and statistically significant genetic differentiation between both countries in the case of the Y-chromosome genetic variation and considering both types of markers. The strong genetic separation of both countries was also revealed in a pairwise analysis of F_{ST} based on Y-SNPs and R_{ST} based on Y-STR haplotypes presented here by MDS plots (Fig. 2a, b); all Polish populations cluster together and are strongly separated from a cluster containing all German populations.

As a further test for geographical population substructure, we performed Monmonier's analysis for detecting the presence of genetic barriers given the spatial distribution of the populations, using Y-SNP-based F_{ST} and Y-STR-based R_{ST} values separately. Based on both datasets, we obtained exactly the same genetic barrier between Poland and Germany (Fig. 3) that we also observed by means of SAMOVA. This means that populations were clustered according to their country of origin by use of the Y-chromosome data.

As also evident from the AMOVA results and the MDS plots, there is a striking genetic homogeneity within Poland based on Y-SNPs and Y-STRs. Only 0.3% for Y-SNPs and 0.08% for Y-STRs of the total genetic variation, both F_{ST} value are not statistically significant, are expressed between Polish populations (Tables 4, 5). In contrast, we observed a small but statistically significant population differentiation within Germany, with 1.4% for Y-SNPs and 1% for Y-STRs (Tables 4, 5). We tested if a grouping of the German populations according to longitude into Eastern

German populations (Rostock, Greifswald, Berlin, Magdeburg, and Leipzig) and Western German populations (Hamburg, Cologne, Muenster, Mainz, Munich, and Freiburg) can explain the observed substructure within Germany. Indeed, an AMOVA considering all German populations revealed a differentiation of Eastern and Western German populations supported by a small but statistically significant amount of 1.0% for Y-SNPs and 1.3% for Y-STRs of the total variation being expressed between these two groups of populations (Tables 4, 5). A clustering of East and West German populations is also evident from the MDS plot of pairwise F_{ST} and R_{ST} distances (Fig. 2a, b) and can be explained by a higher frequency of haplogroup R1*(xR1a1) in populations from Western Germany compared with those from Eastern Germany and vise versa for haplogroup R1a1*. This also explains the placement of all East German populations between West German and Polish populations (although highly separated from the latter) in both MDS plots (Fig. 2a, b).

Correlation of Y-SNP haplogroups and Y-STR haplotypes

Initiated by the correspondence of Y-SNP and Y-STR results in the AMOVA, Barrier and MDS analyses, we performed a Mantel test comparing the genetic distance matrices from the population pairwise Y-SNP-based F_{ST} and the Y-STR-based R_{ST} analysis in order to test for correlation of the Y-SNP haplogroup and the Y-STR haplotype data. As might be expected from the previous results, we obtained a highly statistically significant positive correlation between both genetic distance matrices ($r = 0.959$; $P = 0.001$), which only slightly decreased when controlling for the geographical distance ($r = 0.925$, $P = 0.001$).

Relative contribution of Y-SNPs and Y-STRs to population differentiation

We were interested in the individual contribution of the different haplogroups to the observed population differentiation between Germany and Poland. Therefore, the distributions of the populations according to their Y-SNP haplogroup frequencies were plotted by means of correspondence analysis in a two-dimensional plot (Fig. 4). The first dimension explains 65% of the overall variance and separates clearly the populations according to their country of origin, namely Germany and Poland; the second dimension only explains 9% of the overall variance, thus indicating that the largest differences are due to the division between Germany and Poland. Plotting the haplogroups in the same graphical representation opens the possibility to assess which haplogroups are contributing to the distribution of populations in the plot. Polish populations tend to cluster together, due to the high frequency of the haplogroup R1a1* they contain, although haplogroup N3* also contributes to the separation of the Polish groups. On the other hand, German populations are separated from Polish populations due to the presence of R1*(xR1a1), although other haplogroups occurring in minor frequencies, such as P*(xR1), also have an influence to the distribution of German populations in the dimensional space.

We also performed AMOVA based on Y-STR haplotypes associated with the three most common Y-SNP haplogroups, R1*(xR1a1), R1a1* and I*. We obtained very small but still statistically significant differentiation between German and Polish populations for R1*(xR1a1) ($F_{CT} = 0.02800$, $P = 0.00098$), and also for R1a1* ($F_{CT} = 0.00899$, $P = 0.00978$). However, in the MDS plots from pairwise R_{ST} distances based on Y-STR haplotypes associated with one or the other haplogroup, we could not detect any clustering according to both countries (data not shown). But when we used Y-STR haplotypes associated with both haplogroups we obtained a large and statistically significant differentiation between Germany and Poland ($F_{CT} = 0.28513$, $P < 0.00001$). Also, Polish and German populations are highly differentiated in an MDS plot from pairwise R_{ST} distances considering pooled R1*(xR1a1)/R1a1* Y-STR haplotypes (data not shown) highly similar to the MDS plot using Y-STR data from all haplogroups (Fig. 2b).

Surprisingly, we also observed a large and statistically significant differentiation between Poland and Germany when performing AMOVA for Y-STR haplotypes associated with haplogroup I* ($F_{CT} = 0.14707$, $P < 0.00001$). Also, the MDS plot based on pairwise R_{ST} values revealed a clear separation between all German populations on one side and all Polish populations on the other side (data not shown). Recently, five subgroups of haplogroup I*, identified by additional Y-SNPs, were studied in a large set of mostly European populations (Rootsi et al. 2004). We were interested to know whether the separation of German and Polish groups, as observed here based on Y-STRs associated with haplogroup I*, could be explained by the two different haplogroup I* subgroups. Based on five Y-STRs that were analysed in both studies (DYS19-DYS390-DYS391-DYS392-DYS393), we identified 90 haplotypes among the 287 Germans carrying haplogroup I* (haplotype diversity: 0.9145 ± 0.0126) and 50 haplotypes among the 158 Polish men with haplogroup I* (0.9298 ± 0.0121). Also, pooled German and Polish haplogroup I* samples were significantly different based on R_{ST} using those five Y-STRs ($R_{ST} = 0.12815$, $P < 0.00011$). In Germany, the most frequent haplotype (14-22-10-11-13) occurred in 75 out of 287 (26.1%) individuals, and the second most frequent haplotype (14-23-10-11-13) in 27 out of 287 (9.4%). Both haplotypes differ from each other by one repeat at one Y-STR locus (DYS390), thus they are closely related. These two haplotypes together occur in 102 out of 287 (35.5%) German haplogroup I* individuals. Interestingly, these two haplotypes are also the two most frequent Y-STR

haplotypes associated with haplogroup I* subgroup I1a*(xI1a4, I1b, I1c)—determined by the M253 mutation and occurring in 95 out of 189 (50.3%) hgI1a* individuals for which combined Y-STR/Y-SNP data were available (Rootsi et al. 2004). In our Polish samples, these two haplotypes occurred in 26 out of 158 (16.5%) haplogroup I* individuals. Moreover, when considering the most common German Y-STR haplotype, plus all of its one-repeat step neighboring haplotypes, 131 out of 287 (45.6%) German individuals were covered, as well as 129 out of 189 (68.3%) haplogroup I1a*(xI1a4, I1b, and I1c) individuals (Rootsi et al. 2004), whereas only 39 out of 158 (24.7%) Polish haplogroup I* men in our study.

The difference between German and Polish hgI* Y chromosomes is even more apparent from the Polish perspective. The most frequent Polish haplogroup I* haplotype (16-24-11-11-13) occurred in 33 out of 158 (20.9%) individuals, and the second most frequent haplotype (16-24-10-11-13) in 15 out of 158 (9.5%). Both haplotypes differ from each other by only one repeat at one Y-STR locus (DYS391), suggesting that they are closely related. These two haplotypes together account for 48 out of 158 (30.4%) Polish haplogroup I* individuals. Interestingly, these two haplotypes are the two most frequent Y-STR haplotypes associated with haplogroup I1b*(xI1a, 1a4, and I1c)—determined by the mutation P37 and occur in 116 out of 220 (52.7%) haplogroup I1b individuals for which combined Y-STR/Y-SNP data are available (Rootsi et al. 2004). In our German samples, these two haplotypes occur in only seven out of 287 (2.4%) haplogroup I* individuals. Furthermore, considering the most frequent Polish haplotype plus all one-step neighbors, 66 out of 158 (41.8%) Polish individuals are covered, as well as 166 out of 220 (75.5%) of haplogroup I1b*(xI1a, 1a4, and I1c) individuals (Rootsi et al. 2004), whereas only 19 out of 287 (6.6%) German haplogroup I* men in our study.

Spatial autocorrelation analysis for Y-SNP haplogroups

We have tested by means of spatial autocorrelation analysis the spatial distribution of the Y-SNP haplogroups observed in our dataset. The only haplogroups that tend to show statistically significant clinal patterns are R1a1* and R1*(R1a1). In the case of R1a1*, this clinal pattern decreases from east to west as can be seen by the large correlation observed with longitude ($r = 0.925$, $P < 0.001$); on the other hand, R1*(xR1a1) tends to correlate both with longitude ($r = -0.88$, $P < 0.001$) and with latitude ($r = -0.463$, $P < 0.046$), thus suggesting a west to east clinal pattern (Fig. 5). We also performed this analysis for the two haplogroup I* subgroups I1a and I1b as inferred by Y-STR haplotype analysis and using only those haplogroup I* individuals that carry the two most frequent Y-STR haplotypes associated with each of the two subgroups. We observed clinal patterns for both haplogroups I1a and I1b, east to west in the case of I1a (correlation with longitude $r = -0.809$, $P < 0.0001$) and west to east in the case of I1b (correlation with longitude $r = 0.86$, $P < 0.00001$) (Fig. 5). However, the low autocorrelation level for the first geographic distance class in all of the autocorrelograms analysed (see Fig. 5) should be taken into consideration; this result indicates that the spatial structure we observe is produced by the difference at large geographic distances, but not at smaller ones.

Discussion

While studying the distribution of two types of Y-chromosomal markers, Y-SNPs and Y-STRs, in regional population samples from the present-day territory of Germany and Poland, we found statistically significant differences in the distribution of paternal lineages between both countries. Furthermore, the SAMOVA approach revealed a significant grouping of all population samples analysed into two groups: on one hand, all regional population samples from Germany, and on the other hand, all regional population samples from Poland. Based on AMOVA, we quantified this genetic differentiation and observed a large and statistical significant amount of 15% (Y-SNPs) or 14% (Y-STRs) of the respective total genetic variation being explained by differences between the two countries. This political population differentiation was confirmed by means of the Monmonier's algorithm with an obtained genetic border between Poland and Germany that closely resembles the course of the political border between both countries for both Y-STR and Y-SNP data. Furthermore, we observed a statistically significant Y-SNP/Y-STR homogeneity within Poland, which is underlined by the fact that we could confirm our previous Y-STR results (Ploski et al. 2002) by including here not only two additional Polish populations (Suwalki and Szczecin) but also independent individual samples for the regions used before (except Warsaw, and partly Gdansk and Bydgoszcz). In contrast to the Polish data, Y-chromosome diversity was less homogeneous within Germany and we identified small but statistically significant Y-chromosome differences between Eastern and Western German populations as defined by longitude. This geographical east/west separation also reflects a political separation between 1949 and 1989 due to the two German states that became a human separation between 1961 and 1989.

The Y-SNP data that were generated in the present study—in addition to the Y-STR data—provide evidence on the molecular basis of the observed genetic differentiation as well as contribute to the overall explanation of the observed genetic differences between Poland and Germany. Although a statistically significant differentiation between Poland and Germany was observed when using all Y-SNP haplogroups detected, we demonstrated that this phenomenon was mainly—but not exclusively—caused by two Y-SNP

haplogroups and their associated Y-STR haplotypes: R1*(xR1a1) together with haplogroup R1a1*. Previously, it has been suggested that the M173 A to C mutation, determining haplogroup R1*(xR1a1) originated 40–35,000 y.a. in Western Europe, perhaps the Iberian peninsula, and that the M17 G deletion, determining haplogroup R1a1*, arose (on a M173 Y chromosome) later on in Eastern Europe, e.g. the present-day Ukraine (Semino et al. 2000). Furthermore, it has been argued that both haplogroups expanded into central Europe after the last glacial maximum (20,000–13,000 y.a.) (Semino et al. 2000). When population samples from all over Europe were considered previously a statistically significant clinal frequency distribution of haplogroups R1*(xR1a1) and R1a1* has been observed (Semino et al. 2000; Rosser et al. 2000) with haplogroup R1*(xR1a1) being highly frequent in Western Europe and decreasing in frequency towards Eastern Europe, and vice versa for haplogroup R1a1*, being highly frequent in Eastern Europe and decreasing in frequency towards Western Europe. Those clinal frequency distributions for haplogroup R1a1* and R1*(xR1a1) have been associated with different ancient population movements in Europe and additional clines have been observed for other Y-chromosome haplogroups and were associated with other ancient migration waves (Rosser et al. 2000; Semino et al. 2000).

Although the majority of the genetic heterogeneity between Polish and German populations was caused by differences in the distribution of haplogroups R1a1* and R1*(xR1a1), we also showed that haplogroup I* individuals contributed to the phenomenon, albeit to a smaller degree given the lower frequency observed. On the basis of differences in the Y-STR distribution in Polish and German males with haplogroup I* using recently published data (Rootsi et al. 2004), we found indirect evidence that the most prevalent subtype of haplogroup I* in Poland is I1b*, whereas in Germany it is I1a*(xI1a4, I1b, I1c). Previously, it has been suggested that haplogroup I1a*(xI1a4, I1b, I1c) originated in Western Europe (Rootsi et al. 2004) and it was previously found more than four-times more frequently in Germany (25%) than in Poland (5.8%). Furthermore, haplogroup I1a* shows a clinal frequency distribution across Europe with high frequencies in Northwest Europe to low frequency in Southeast Europe (Rootsi et al. 2004). This agrees with our observation of a more than twofold higher frequency of the two most common haplogroup I1a*-associated Y-STR haplotypes in our German sample compared with our Polish sample, or about twofold higher when considering all one-step neighboring haplotypes. On the other hand, it has been argued elsewhere that haplogroup I1b*(xI1a, I1a4, and I1c) originated in Eastern Europe and was previously found ten-times more frequently in Poland (9.9%) than in Germany (0%) (Rootsi et al. 2004). This agrees with our observation of a more than 12-times higher frequency of the two most common Y-STR haplotypes associated with haplogroup I1b in our Polish samples compared with our German sample, or more than six-times higher considering all one-step neighboring haplotypes. Therefore we can assume that the statistically significant difference in haplogroup I* between Germany and Poland as detected here using Y-STR haplotypes is—at least to a large degree—caused by differences in the distributions of the two haplogroup I* subgroups, I1a and I1b, and their associated Y-STR haplotypes.

The question appears why we see a strong and statistically significant differentiation for haplogroup R1*(xR1a1) together with R1a1*, and also for the two inferred haplogroup I* subgroups, I1a and I1b, between regional populations from the geographically neighboring countries Germany and Poland, although clinal frequency distributions—explained by ancient population movements—have been previously observed for these four haplogroups across Europe (Rosser et al. 2000; Semino et al. 2000; Rootsi et al. 2004)? When we performed spatial autocorrelation analysis to test for statistically significant clinal frequency distribution, we observed for all four haplogroups an autocorrelogram compatible with a clinal pattern except in the case of the first geographical distance class, which shows a lower autocorrelation than expected in a clinal pattern. This indicates that the spatial structure we observe is produced by the difference at large distances but not at smaller ones, which can be explained by the presence of the genetic barrier that we have detected by means of SAMOVA and Monmonier's algorithm. The first geographic distance class contains mainly the pairs of populations on the same side of the barrier (which tends to lead autocorrelation values close to 0 due to the homogeneous pattern), whereas for larger geographic distance classes the pairs of populations correspond mainly to one population on each side of the barrier (thus recreating the ancestral clinal pattern). Taking into account that the clinal frequency distributions of the Y-SNP haplogroups across entire Europe are mainly explained by ancient population movements (Rosser et al. 2000; Semino et al. 2000; Rootsi et al. 2004), the presence of the genetic barrier that we have detected has to be established after the creation of these clinal patterns (otherwise the barrier would have prevented the establishment of the clines). In addition, the strong positive correlation between Y-SNP and Y-STR data, as observed here, implies that the reason for the genetic population differentiation must be recent; otherwise, the relatively high mutation rate of Y-STRs (Kayser et al. 2000) would tend to destroy the correlation. Since the genetic barrier we observe superimposes to the actual political borders between Germany and Poland, which was established shortly after the Second World War (WWII), we suggest here that our observation of statistically significant genetic differentiation between Poland and Germany, as well as genetic homogeneity within Poland, could be explained by the severe human resettlements during and shortly after WWII and thus

the formation of the present-day Polish and German states.

The present-day Polish/German territory has experienced very recent and severe population movements as a consequence of WWII. It is estimated that during 1944 and 1951 more than eight-million people of German origin—which inhabited the territory of present-day Poland for hundreds of years (e.g. East Prussia, Silesia, and Pomerania)—moved westwards into the present-day Germany either escaping the advancing eastern front-line of WWII or due to the politically forced resettlements shortly after WWII (Encyclopædia Britannica 2005; Nowa Encyklopedia Powszechna PWN 2004). At the same time, approximately five-million people of mostly Polish descent were forced to move from the region of present-day Ukraine, Lithuania, Belarus, and partly Russia into the present-day Polish territory, whereas half-a-million people previously living in Poland moved into the opposite direction between 1939 and 1944 (Encyclopædia Britannica 2005; Nowa Encyklopedia Powszechna PWN 2004). These numbers constitute a significant proportion of the 28-million people of multiple origin that were living in the present-day territory of Poland before WWII and thus it is likely that the forced migrations associated with WWII in an exceptional way distorted the genetic landscape of the region shaped over the ages by 'natural' demographic processes. Furthermore, these forced movements were restricted by the establishment of the present-day political border between the two states Poland and Germany immediately after WWII. This political border became the border for forced migration of millions of Germans that were resettled to the west of this border and millions of Polish and other people of Eastern European origin that were resettled to the east of this border. The Y-chromosome data presented here suggest that these processes led to a shift of central/western European Y chromosomes characterised by a high frequency of haplogroup R1*(xR1a1) (and less frequent I1a*) towards the west into present-day Germany and, shortly after, a shift of Eastern European Y chromosomes characterised by a high frequency of haplogroup R1a1* (and less frequent I1b*) towards the west into present-day Poland. This recent process of "nation building" of Germany and Poland based on shared cultural (language, religion, and tradition) identities during and immediately after WWII stopped at the present-day political border between both countries that was assigned after WWII, which clearly reflects—as we show here—a statistically significant genetic border in the distribution of human male lineages in this part of Europe. Our genetic data also imply that at least male genetic admixture between people of German and those of Polish origin during the hundreds of years before WWII where they shared the same territory must have been small, as we discussed elsewhere (Ploski et al. 2002).

We also observed statistically significant Y-chromosome differences based on Y-SNPs and Y-STRs within Germany, namely between Eastern and Western German populations. Also, in all MDS plots using either combined Y-SNP, or combined Y-STR data (Fig. 2a, b), or haplogroup R1*(xR1a1)/R1a1*-associated Y-STR data (data not shown) Eastern German groups appeared always clustered together and somewhat separated from Western German groups and their location is always between Western German groups on one side and Polish groups on the other side (but still highly separated from the latter). This can be explained by a higher frequency of haplogroup R1a1* in Eastern (24.3%) than in Western Germany (12.7%) but a lower frequency of haplogroup R1*(xR1a1) in Eastern (34.7%) than in Western Germany (42.4%), and the distribution of respectively associated Y-STR haplotypes. Frequency differences between Eastern and Western German groups are approaching significance for haplogroup R1*(xR1a1) (Mann–Whitney *U*-test: $Z = -1.826$, $P = 0.068$) and are statistically significant for haplogroup R1a1* (Mann–Whitney *U*-test: $Z = -2.739$, $P = 0.006$). This East-West/West-East scenario observed within Germany is somewhat similar to the overall picture we observed between Germany and Poland, but much less pronounced. No statistically significant differentiation in the pairwise F_{ST}/R_{ST} analysis was detected between East and West German populations, whereas almost all pairwise comparisons between German and Polish groups revealed statistically significant differences (Fig. 2a, b). We therefore conclude that Y-chromosome differences between Eastern and Western Germany might be more likely due to more ancient events in the history of European populations, namely a higher eastern European (i.e. Slavic) influence in Eastern (but less in Western) Germany and the higher western European influence in Western (but less in Eastern) Germany. A strong Slavic influence on today's Eastern German territory is well documented, e.g. by the Slavic names of many villages or towns that are not found in Western Germany or by the higher frequency of surnames with Slavic origin in Eastern Germany compared with Western Germany.

Our observation of statistically significant population substructure in closely neighboring areas in Europe has also practical consequences for the forensic application of Y-chromosome markers in Europe. Over the last decade Y-chromosome DNA analysis became successfully established and is now widely used in forensic genetics for the identification of male-specific genetic material, e.g. from rape and sexual assault cases (Kayser 2003). Due to the hypervariability of Y-STR-based haplotypes, innocent suspects (and their paternal lineages) can be excluded with a high degree of accuracy. However, when a match is found, Y-STR haplotype frequencies are needed in order to calculate match probabilities for which Y-STR frequency databases have been started to become established (Roewer et al. 2001; Kayser et al. 2002; Lessig et al. 2003). The largest Y-STR haplotype database publicly available is the Y-chromosome Haplotype Reference Database—YHRD (http://www.yhrd.org), which—as of

March 2005—comprised 28,650 haplotypes in a set of 249 worldwide populations, of which 17,373 haplotypes are from 126 European populations (including to a large degree Y-STR data from this study). This database allows haplotype frequency search and provides frequency estimates for regional populations, but also based on pooled population data. Our results, which have clearly identified population substructure based on both Y-STR haplotypes as well as Y-SNP haplogroups in two neighboring European countries, strongly suggest either the use of regional databases for frequency estimation or (better) the use of more global databases which take into account information on regional population substructure. Activities are currently underway to make knowledge on population substructure available for the YHRD by implementing the recently identified population clusters within Europe (Eastern, Southeastern, Central/Northern and Western Europe as well as Finland) and offering Y-STR haplotype frequency estimates separately for such metapopulations (Roewer et al. 2005).

On the other hand, evidence for strong genetic homogeneity within larger geographic regions, e.g. as observed here for Poland, provides important information for association mapping for disease (and other) gene identification. However, it should be noted here that the evidence we provide in the present study comes from one genetic locus (the Y chromosome) and autosomal genetic evidence needs to be established as well.

Finally, we would like to emphasise what enabled us to detect the genetic signature of an event in human population history as recent as about 50 years ago. We believe that this was possible because of a combination of at least five genetic or non-genetic components: (1) the large number of many millions of individuals for each of the two groups that moved in a relatively short time period (a few years); (2) that the two groups moved discontinuously due to the establishment of a new political border, which therefore became the border for migration; (3) that the two groups on the move were originally characterised by two different high-frequency Y-chromosome SNP markers (and their associated Y-STR haplotypes), which were also used for detection; (4) that the frequency distribution of both Y-SNP markers was originally clinal (but in the opposite directions) due to ancient population movements; and (5) that the border of migration for both groups on the move was perpendicular to the direction of the previously established frequency clines of both Y-SNP markers. An additional influence might come from the non-recombining inheritance of the Y-chromosome markers analysed, although this effect should be much smaller due to the small number of generations that have passed since this recent event.

Acknowledgements We are grateful to Anett Bagans, Silke Brauer, Anita Dettlaff-Kakol, Dennis Kappei, Carmen Krüger, Anja Manek, Nicola Meier, Regina Paszkowska, Monika Reichert, Kerstin Richter, Anke Riedel, and Katja Tremel for expert technical assistance. Richard Cordaux and Vano Nasidze are acknowledged for sharing PCR protocols prior to their original publication. M. Kayser is grateful to Mark Stoneking for providing laboratory facilities at MPI-EVA as well as for his general support, and thanks Gunter Weiss for statistical advice. M. Wozniak acknowledges the support by the Ludwik Rydygier University School of Medical Sciences in Bydgoszcz (grant BW 87/2003). R. Ploski, G. Bargel, and P. Kuzniar all acknowledge the support by the Polish State Committee for Scientific Research (grant 0T00C00322), and by the Medical University of Warsaw (grant 1MB/N/2004). M. Kayser and O. Lao acknowledge the support by the Erasmus University Medical Center Rotterdam and the Netherlands Forensic Institute.

References

Barbujani GV (2000) Geographic patterns: how to identify them and why. Hum Biol 72:133–153

Barbujani G, Goldstein DB (2004) Africans and Asians abroad: genetic diversity in Europe. Annu Rev Genomics Hum Genet 5:119–150

Bender K, Stradmann-Bellinghausen B, Rittner C, Schneider PM (2003) Comparative analysis of STRs and SNPs on the Y-chromosome in Germans, Chinese and Thais. Legal Med 5:164–168

Bosch E, Calafell F, Rosser ZH, Nørby S, Lynnerup N, Hurles ME, Jobling MA (2003) High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. Hum Genet 112:353–363

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368:455–457

Capelli C, Redhead N, Abernethy JK, Gratrix F, Wilson JF, Moen T, Hervig T, Richards M, Stumpf MPH, Underhill PA, Bradshaw P, Shaha A, Thomas MG, Bradman N, Goldstein DB (2003) A Y chromosome census of the British Isles. Curr Biol 13:979–984

Cordaux R, Aunger R, Bentley G, Nasidze I, Sirajuddin SM, Stoneking M (2004) Independent origins of Indian caste and tribal paternal lineages. Curr Biol 14:231–235

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

Dupanloup I, Schneider S, Excoffier L (2002) A simulated annealing approach to define the genetic structure of populations. Mol Ecol 11:2571–2571

Encyclopædia Britannica (2005) History of Poland in Encyclopædia Britannica online (http://search.eb.com/eb/article?tocId = 28216)

Hammer MF (1994) A recent insertion of an Alu element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Hammer MF, Horai S (1995) Y chromosomal DNA variation and the peopling of Japan. Am J Hum Genet 56:951–962

Hurles ME, Maund E, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA (2003) Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. Am J Hum Genet 72:1282–1287

Jin L, Su B (2000) Natives or immigrants: modern human origin in East Asia. Nat Rev Genet 1:126–133

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

Kayser M (2003) The human Y chromosome—introduction into genetics and applications. Forensic Sci Rev 15:77–89

Kayser M, Brauer S, Weiss G, Underhill P, Roewer L, Schiefenhövel W, Stoneking M (2000a) Melanesian origin of polynesian Y chromosomes. Curr Biol 10:1237–1246

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 (2000b) 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, Schiefenhövel W, Underhill P, Stoneking M (2001) Independent histories of Y chromosomes from Melanesia and Australia. Am J Hum Genet 68:173–190

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 US populations. J Forensic Sci 47:513–519

Kayser M, Brauer S, Weiss G, Schiefenhövel W, Underhill P, Shen P, Oefner P, Tommaseo-Ponzetta M, Stoneking M (2003) Reduced Y-chromosome, but not mtDNA, diversity in human populations from West New Guinea. Am J Hum Genet 72:281–302

Ke X, Collins A, Ye S (2001a) PIRA PCR designer for restriction analysis of single nucleotide polymorphisms. Bioinformatics 17:838–839

Ke Y, Su B, Song X, Lu D, Chen L, Li H, Qi C, Marzuki S, Deka R, Underhill P, Xiao C, Shriver M, Lell J, Wallace D, Wells RS, Seielstad M, Oefner P, Zhu D, Jin J, Huang W, Chakraborty R, Chen Z, Jin L (2001b) African origin of modern humans in east Asia: a tale of 12,000 Y chromosomes. Science 292:1151–1152

Lessig R, Willuweit S, Krawczak M, Wu F-C, Pu C-E, Kim W, Henke L, Henke J, Miranda J, Hidding M, Benecke M, Schmitt C, Magno M, Calacal G, Delfin FC, de Ungria MCA, Elias S, Augustin C, Tun Z, Honda K, Kayser M, Gusmao L, Amorim A, Alves C, Hou Y, Keyser C, Ludes B, Klintschar M, Immel UD, Reichenpfader B, Zaharova B, Roewer L (2003) Asian online Y-STR haplotype reference database. Legal Med 5:S160–S163

Lessig R, Zoledziewska M, Fahr K, Edelmann J, Kostrzewa M, Dobosz T, Kleemann W (2005) Y-SNP-genotyping—a new approach in forensic analysis. Forensic Sci Int (in press), available online from the Forensic Sci Int website doi:10.1016/j.forsciint.2004.09.129

Manni FC, Heyer GE (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by ''Monmonier's algorithm''. Hum Biol 76:173–190

Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220

Nasidze I, Ling ES, Quinque D, Dupanloup I, Cordaux R, Rychkov S, Naumova O, Zhukova O, Sarraf-Zadegan N, Naderi GA, Asgary S, Sardas S, Farhud DD, Sarkisian T, Asadov C, Kerimov A, Stoneking MM (2004) Mitochondrial DNA and Y-chromosome variation in the Caucasus. Ann Hum Genet 68:205–221

Nowa Encyklopedia Powszechna PWN (2004) Polska. Historia. Wydawnictwo Naukowe PWN SA http://encyklopedia.pwn.pl/58502_1.html

Ploski R, Wozniak M, Pawlowski 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

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, 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, Liebherr 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

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

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic 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 P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic 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

Rosenberg M (2001) Pattern analysis, spatial statistics, and geographic exegesis. Version 1.1, 1.1 edn. Department of Biology, Arizona State University, Tempe

Rosser ZH, Zerjal T, Hurles ME, Adojaan M, Alavantic D, Amorim A, Amos W, Armenteros M, Arroyo E, Barbujani G, Beckamnn 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, Kravchenco 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

Schneider S, Roessli D, Excoffier L (2000) Arlequin Version 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Geneva

Schurr TG (2004) The peopling of the New World: perspectives from molecular Anthropology. Annu Rev Anthropol 33:551–583

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

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, Underhill PA (2002) Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. Am J Hum Genet 70:265–268

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

Soodyall H, Nebel A, Morar B, Jenkins T (2003) Genealogy and genes: tracing the founding fathers of Tristan da Cunha. Eur J Hum Genet 11:705–709

Stoneking M, Soodyall H (1996) Human evolution and the mitochondrial genome. Curr Opin Genet Dev 6:731–736

Thomson R, Pritchard JK, Shen P, Oefner PJ, Feldman MW (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. Proc Natl Acad Sci USA 97:7360–7365

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ (2000) Y-chromosome sequence variation and the history of human populations. Nat Genet 26:358–361

Weale ME, Weiss DA, Jager RF, Bradman N, Thomas M (2002) Y chromosome evidence for Anglo-Saxon mass migration. Mol Bio 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 USA 98:10244–10249

Wilson JF, Weiss DA, Richards M, Thomas MG, Bradman N, Goldstein DB (2001) Genetic evidence for different male and female roles during cultural transitions in the British Isles. Proc Natl Acad Sci USA 98:5078–5083

Zegura SL, Karafet TM, Zhivotovsky LA, Hammer MF (2004) High-resolution SNPs and microsatellite haplotypes point to a single, recent entry of Native American Y chromosomes into the Americas. Mol Biol Evol 21:164–175

Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos F, Schiefenhövel W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjidmaa D, Sajantila A, Salo P, Crawford MH, Evgrafov O, Tyler-Smith C (1997) Genetic relationships of Asians and northern Europeans revealed by Y-chromosomal DNA analysis. Am J Hum Genet 60:1174–1183

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. Am J Hum Genet 72:717–721