

ARTICLE

# Y-chromosomal STR haplotype analysis reveals surname-associated strata in the East-German population

Uta-Dorothee Immel\*<sup>1</sup>, Michael Krawczak<sup>2</sup>, Jürgen Udolph<sup>3</sup>, Angela Richter<sup>4</sup>, Heike Rodig<sup>5</sup>, Manfred Kleiber<sup>1</sup> and Michael Klintschar<sup>1</sup>

<sup>1</sup>Department of Legal Medicine, Martin-Luther-University, Halle (Saale), Germany; <sup>2</sup>Institute of Medical Informatics and Statistics, Christian-Albrechts-University, Kiel, Germany; <sup>3</sup>Institute for Slavistics, University of Leipzig, Leipzig, Germany; <sup>4</sup>Institute for Slavistics, Martin-Luther-University, Halle (Saale), Germany; <sup>5</sup>Biotype AG, Dresden, Germany

In human populations, the correct historical interpretation of a genetic structure is often hampered by an almost inherent inability to differentiate between ancient and more recent influences upon extant gene pools. One method to trace recent population movements is the analysis of surnames, which, at least in Central Europe, can be thought of as traits 'linked' to the Y chromosome. Illegitimacy, extramarital birth and changes of surnames may have substantially obscured this linkage. In order to assess the actual extent of correlation between surnames and Y-chromosomal haplotypes in Central Europe, we typed Y-chromosomal short tandem repeat markers in 419 German males from Halle. These individuals were subdivided into three groups according to the origin of their respective surname, namely German (G), Slavic (S) or 'Mixed' (M). The distribution of the haplotypes was compared by Analysis of Molecular Variance. While the M group was indistinguishable from group G ( $\Phi_{ST} = -0.0008$ ,  $P > 0.5$ ), a highly significant difference ( $\Phi_{ST} = 0.0277$ ,  $P < 0.001$ ) was observed between the S group and the combined G + M group. This surprisingly strong differentiation is comparable to that of European populations of much larger geographic and linguistic difference. In view of the major migration from Slavic countries into Germany in the 19th century, it appears likely that the observed concurrence of Slavic surnames and Y chromosomes is of a recent rather than an early origin. Our results suggest that surnames may provide a simple means to stratify, and thereby to render more efficient, Y-chromosomal analyses of Central Europeans that target more ancient events.

*European Journal of Human Genetics* (2006) 14, 577–582. doi:10.1038/sj.ejhg.5201572; published online 25 January 2006

**Keywords:** Y chromosome; onomastics; surnames; Germans; Slavs

## Introduction

The study of surnames, also called 'the poor man's population genetics',<sup>1</sup> has a long tradition in genealogical research and was extensively used long before the term

'genetics' was first coined.<sup>2</sup> Indeed, the fact that surnames are patrilineally inherited in many parts of the world,<sup>3–5</sup> including Central Europe,<sup>2</sup> implies that names should be of considerable interest to geneticists too. Surnames in combination with genetic studies have proved useful for describing population structures, for example, in France, Sicily and Netherlands.<sup>6–8</sup>

However, Y-chromosomal DNA polymorphisms are ideally suited for studies of male demography themselves, and the availability of rapidly evolving markers on the

\*Correspondence: U-D Immel, Department of Legal Medicine, Martin-Luther-University Halle, Franzosenweg 1, Halle (Saale) 06112, Germany. Tel/Fax: +49 345 557 1587; E-mail: uta.immel@medizin.uni-halle.de  
Received 14 June 2005; revised 10 November 2005; accepted 8 December 2005; published online 25 January 2006

Y chromosome has lately rendered the onomastics of human surnames an outsider discipline. Thus, the ability of hypervariable Y-chromosomal short tandem repeats (Y-STRs) to discriminate even between closely related and co-localized male populations has been demonstrated for Germans and Dutch<sup>9</sup> for the Baltic populations,<sup>10</sup> for Central England and North Wales,<sup>11</sup> and for Poland and Germany.<sup>12</sup> At an even larger scale, the recent identification of previously unrecognized population strata in the Y-STR haplotype distribution of more than 12 000 males from 91 European localities<sup>13</sup> has once more highlighted the usefulness of this approach. Nevertheless, only a small number of studies have so far addressed the actual relationship between the distribution of surnames and Y-chromosomal haplotypes.<sup>14–17</sup>

Surnames vary substantially both between and within European countries. In Germany, for example, although the majority of the one million different surnames are typically German (eg 'Müller', 'Schmidt' or 'Berger'), names with foreign roots are also abundant. The majority of the latter are of Slavic origin<sup>18</sup> (approximately 20%) and many of them are easy to recognize by consonant combinations that are otherwise unfamiliar to the German language. Examples include the names of the German writer Kurt Tucholsky and of the second author of this article. In many cases, however, the foreign origin of a surname may not be immediately apparent, as is the case, for example, for the name of the 18th Century play writer Gotthold Ephraim Lessing.

The patrilineal inheritance of both surnames and Y chromosomes suggests that different strata of surnames should correspond to different strata of Y chromosomes. Since this relationship is likely to have become obscured not only by mutation but also by illegitimate births and the change of surnames, quantifying the residual correlation between the two characteristics would be of both theoretical and practical relevance. On the one hand, information about the history of patrilines is useful for the precise estimation of mutation rates and for the assessment of migration behaviours. On the other hand, surnames potentially provide a simple means of stratifying populations prior to Y-chromosomal analyses that target prehistoric events, thereby increasing their efficiency through a reduction in genotyping load. The aim of the present study was thus to assess the extent to which Y-STR haplotypes of German males, born and living in the region of Halle (Saale), are indicative of a German, Slavic or mixed German-Slavic descent of their surnames.

## Materials and methods

### DNA samples

DNA samples were obtained from 419 German males, born around Halle (Saale), located in the South-East of Germany (Figure 1), who identified themselves as Germans. An

additional group of 29 German males were sampled from the Sorbish minority, a Slavic-speaking community living in the Lausitz area near the Polish border.<sup>19</sup>

### Y-STR analysis

Eight Y-STR loci were analysed, namely DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392 and DYS393. Locus information and PCR primer sequences can be found in Kayser *et al*<sup>20</sup> or at the Y-STR Haplotype Reference Database (YHRD) web site ([www.yhrd.org](http://www.yhrd.org)). The YHRD nomenclature was used here in accordance with recommendations by the International Society of Forensic Genetics,<sup>18</sup> designating Y-STR alleles by the number of repeats included. DNA was amplified in two multiplex reactions, following Elmoznino and Prinz.<sup>21</sup> Consistent allele designation and genotyping quality were assured by the concurrent electrophoretic analysis of sequenced allelic ladders or sequenced reference DNA samples. PCR products were analysed by capillary electrophoresis using an ABI 310 Genetic Analyzer (Applied Biosystems, Weiterstadt, Germany) and the Genotyper software.<sup>22</sup> DNA of the Sorbish males was analysed using the Mentype<sup>®</sup> Argus Y-MH PCR Amplification Kit (Biotype, Dresden).

### Analysis of surnames

The Halle samples were divided into three subgroups, according to surname. Two larger groups comprised 195 males with surnames that were definitely German ('G') and 185 males with definitely Slavic surnames ('S'). The third group contained 39 males with mixed German-Slavic



**Figure 1** Map of Germany showing the city of Halle (Saale).

surnames ('M'). Samples of 29 Sorbs<sup>19</sup> and some 1313 published haplotypes from Polish males<sup>13</sup> were used for comparison. Surname groups were defined on the basis of spelling, using certain combinations of consonants and surname suffixes to categorize the origin of the name in question. Suffixes '-er', '-mann' and '-burg', for example, are typically German whereas '-ke', '-ka', '-ow' and '-ski' are typically Slavic. In addition, the root morphemes of surnames were also examined. Examples for a Slavic root comprise 'Lessing', which sounds German but was derived from the Slavic expression for 'forest settler', and 'Kafka', which in Czech means 'jackdaw'. Mixed surnames include both German and Slavic elements, that is, a German basis and a Slavic ending, or *vice versa* ('Wudtke' or 'Kuppke'). These surnames are the result of a long parallel usage of both German and Slavic languages in the eastern part of Germany.

### Statistical analysis

The genetic relationship between the German, Sorbish and Polish samples was assessed by Analysis of Molecular Variance (AMOVA) using  $\Phi_{ST}$ , an analogue of Wright's  $F_{ST}$  that takes the evolutionary distance between individual Y-STR haplotypes into account.<sup>23,24</sup> The analysis was confined to the so-called 'core' haplotype, comprising all markers but DYS385. Marker DYS385 had to be excluded since its multilocal nature hampers the unambiguous assignment of evolutionary distances to allele pairs. Populations were recursively clustered by combining, in each step, that pair of samples or clusters that yielded the minimum global  $\Phi_{ST}$  value for the core haplotype. Clustering was carried out until only one cluster remained. Estimates of pairwise and global  $\Phi_{ST}$  values were obtained using the ARLEQUIN software<sup>25</sup> with a single step mutation model, and tested for statistical significance by means of random permutation of samples in 10 000 replicates.<sup>13</sup>

### Results

In the 419 East-German males analysed in the present study, a total of 270 different Y-STR haplotypes were observed. While the most frequent haplotype occurred 10 times, 146 haplotypes were unique (data available from the authors upon request). Group G comprised 139 different surnames, 18 of which occurred twice. Five surnames were observed more than two times ( $3 \times 3$ ,  $1 \times 5$ ,  $1 \times 6$ ). There was only one instance in group G of a surname being shared by two males with the same haplotype. In group S, 177 different surnames occurred, four of which were found twice. No two males with the same surname had the same haplotype. Finally, no shared surnames and haplotypes were observed in group M.

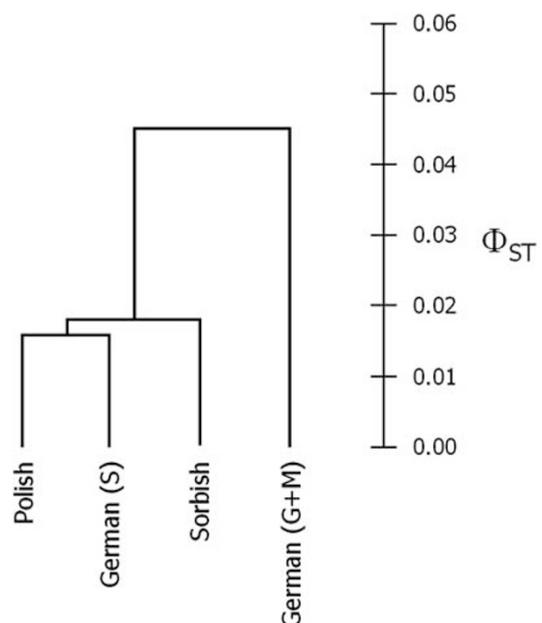
Upon AMOVA, the core Y-STR haplotype distributions of males with German ('G') and mixed surnames ('M') were found to be indistinguishable ( $\Phi_{ST} = -0.0008$ ,  $P > 0.5$ ). The

two samples were therefore combined into one group ('G + M'). Please note that this joint consideration of G and M was retrospectively justified in that an analysis of group G alone yielded virtually identical results (not shown). A highly significant difference emerged between the combined G + M group and the group of males with a Slavic surname ('S'; see Table 1). The observed level of differentiation (pairwise  $\Phi_{ST} = 0.0277$ ,  $P < 0.001$ ) between groups G + M and S was surprisingly large and so were approximates seen between European populations of much larger geographical and linguistic distances (eg Cologne and Budapest; see [www.yhrd.org](http://www.yhrd.org)). Cluster analysis based upon global  $\Phi_{ST}$  (Figure 2) revealed that the Y-STR core haplotype distribution of the German S group is substantially closer to that of the Polish population than to that of the G + M group. The Sorbish males appear to be similarly close to both the S group and the Polish group, although their positioning in the tree may be less robust owing to small sample size.

**Table 1** Pairwise Y-STR-based  $\Phi_{ST}$  for central European males

Origin	German (S)	Sorbish	Polish
German (G+M)	0.0277	0.1202	0.0711
German (S)		0.0437*	0.0159
Sorbish			0.0259**

All  $\Phi_{ST}$  values are significantly larger than zero (randomization  $P < 0.001$  except \* $P = 0.013$ , \*\* $P = 0.002$ ).



**Figure 2** Clustering of Central European male samples by global Y-STR-based  $\Phi_{ST}$ .

In a recent study of European Y-STR haplotypes, several population clusters were identified; among them were clearly defined 'Eastern European' and 'Western European' groupings.<sup>13</sup> Haplotypes from these fringe clusters, as well as their one-step neighbours, were classified as either 'Western' or 'Eastern', depending upon where they were more frequent. A similar characterization of the present samples in terms of the relative proportion of the fringe haplotypes resulted in highly significant differences between the two surname-defined German subgroups, G + M and S ( $\chi^2 = 13.094$ , 2 df,  $P = 0.001$ ). While 88 of the 234 haplotypes (38%) in the combined G + M group were classified as 'Western', this was the case for only 42 of the 185 haplotypes (23%) in group S. In contrast, 80 G + M haplotypes (34%) were of 'Eastern' type compared to 91 S haplotypes (49%). The portion of unclassifiable haplotypes was 28% in both groups (66 in G + M, 52 in S).

The seeming characterization of surname-defined male samples from Halle as either 'Western' or 'Eastern' was further corroborated by comparing the frequency of all haplotypes observed in groups G + M and S with the current release of YHRD (Release 15), comprising 17 214 haplotypes from 125 samples of European or Near-Eastern extraction (Figure 3). Males from group G + M shared the majority of their Y-STR haplotypes with western populations whereas the distribution in group S was closer to that of eastern, most notably Polish, populations. The proportion of haplotypes shared between group S and Polish males was higher than that with any other German sample.

## Discussion

How can the profound stratification observed among East-German male lineages and their correlation with surnames be best explained? Although the name 'Germany' appears to imply a homogenous origin of the German people, the country has always been a gateway for migration, mostly from east to west. The best documented wave of migration was that of Eastern Germanic tribes and Slavs, driven by the Huns, that led to the downfall of the Roman Empire. In historic times, two major instances of assimilation of Slavic people into the German nation occurred. Around 950 AD, the German Empire started to put pressure upon the Slavic peoples inhabiting large areas of what was to become, in the mid of the 20th Century, the German Democratic Republic.<sup>26</sup> By 1100 AD, after more than 100 years of wars and proselytization, the complete area of contemporary Germany had come under the influence of the German Empire. During the following centuries, most of the non-Germanic tribes (like the Baltic Prussians) completely abandoned their language, and their descendants are today regarded as 'typically German'. Only in a small area, southeast of Berlin, known as the Lausitz, the Slavic-

speaking Sorb people maintained their language and culture, and their descendants today represent the only recognized, non-immigrant minority in East Germany. In any case, the names of many cities, including Berlin (meaning 'little swamp'), and some surnames, most notably those of 'typically Prussian' nature like 'von Clausewitz' or 'Virchow', still reflect the Slavic roots of this part of Germany. The second major assimilation of people with Slavic ancestry occurred during the Industrial Revolution in the 19th Century. Thousands of people from Eastern Europe migrated to the West to work in the surging industrial areas of Germany (Silesia, Ruhr-Area). Although they brought their surnames with them, they nevertheless became culturally amalgamated quite rapidly by the German majority.

The Halle region is located exactly at the intersection of the Germanic and Slavic spheres of influence of the 10th century, but it is also a traditional mining and chemical industry area (Halle-Leipzig-Bitterfeld) that has attracted Slavic workers during the Industrial Revolution. Both of these factors should have had an impact upon the male-specific genetic structure of the local population where surnames of Germanic and Slavic origin are about equally frequent. In terms of the relative importance of the two historic instances for the observed correlation between Y-STR haplotypes and surname characteristics, it is interesting to note that surnames first occurred in Europe in Venice during the 9th Century. From there, the law of name bearing was adopted in France and Catalonia in the 11th, and in England, and Western and Southern Germany in the 12th Century. In the North and East of Germany, the custom was practised no earlier than the 15th Century and, in some rural regions, surnames became fashionable only in the 18th century, nearly 900 years after their first appearance in Europe.<sup>27,28</sup> Furthermore, surnames frequently changed or became modified until the beginning of the 19th century. Therefore, it appears unlikely that the correlation between surnames and Y-STR haplotypes observed in our study dates back to the Middle Ages, but is more likely to be the result of the immigration of industrial workers in the 19th Century instead. In this respect, Central Europe appears to differ from England and Ireland where patrilineally inherited names are presumed to have a much deeper rooting.<sup>14-17</sup>

Our results highlight the fact that the Y-chromosomal genetic structure of modern Central European populations is heterogeneous and that, particularly in East Germany, the concomitant strata may be resolvable by the consideration of surnames. This implies that future studies targeted at more ancient population movements inside or outside the region through the use of slowly evolving Y-chromosomal markers (ie SNPs) may gain efficiency from allotting the genotyping load according to surnames.



**Figure 3** Matches between the Y-STR Haplotype Reference Database and core Y-STR haplotypes of males with German or Mixed (top) and Slavic (bottom) surnames.

**Acknowledgements**

We thank Tim Lu for helpful comments on the manuscript and Gerald Bothe for graphical work.

**References**

- 1 Crow JF: Surnames as markers of inbreeding and migration. Discussion. *Hum Biol* 1983; 55: 383–397.
- 2 Jobling MA: In the name of the father: surnames and genetics. *Trends Genet* 2001; 17: 353–357.
- 3 Bach A: *Deutsche Namenkunde. Die deutschen Personennamen in geschichtlicher, geographischer, soziologischer und psychologischer Betrachtung*. Heidelberg: Bach A, 1953.
- 4 Gottschald M: *Deutsche Namenkunde*. Berlin, New York: de Gruyter, 1982.
- 5 Heintze A, Cascorbi P: *Die deutschen Familiennamen*. Halle: Georg Olms Verlag AG, 1933.

- 6 Degioanni A, Darlu P, Raffoux C: Analysis of the French national registry of unrelated bone marrow donors, using surnames as a tool for improving geographical localisation of HLA haplotypes. *Eur J Hum Genet* 2003; **11**: 794–801.
- 7 Zei G, Lisa A, Fiorani O *et al*: From surnames to the history of Y chromosomes: the Sardinian population as a paradigm. *Eur J Hum Genet* 2003; **11**: 802–807.
- 8 Manni F, Toupance B, Sabbagh A, Heyer E: New method for surname studies of ancient patrilineal population structures, and possible application to improvement of Y-chromosome sampling. *Am J Phys Anthropol* 2005; **126**: 214–228.
- 9 Roewer L, Kayser M, Dieltjes P *et al*: Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum Mol Genet* 1996; **5**: 1029–1033.
- 10 Lessig R, Edelmann J, Krawczak M: Population genetics of Y-chromosomal microsatellites in Baltic males. *Forensic Sci Int* 2001; **118**: 153–157.
- 11 Weale ME, Weiss DA, Jager RF, Bradman N, Thomas MG: Y chromosome evidence for Anglo-Saxon mass migration. *Mol Biol Evol* 2002; **19**: 1008–1021.
- 12 Ploski R, Wozniak M, Pawlowski R *et al*: Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Hum Genet* 2002; **110**: 592–600.
- 13 Roewer L, Croucher PJ, Willuweit S *et al*: Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet* 2005; **116**: 279–291.
- 14 Sykes B, Irven C: Surnames and the Y chromosome. *Am J Hum Genet* 2000; **66**: 1417–1419.
- 15 Rosser ZH, Zerjal T, Hurler ME *et al*: Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am J Hum Genet* 2000; **67**: 1526–1543.
- 16 Helgason A, Sigurðardóttir S, Nicholson J *et al*: Estimating Scandinavian and Gaelic ancestry in the male settlers of Iceland. *Am J Hum Genet* 2000; **67**: 697–717.
- 17 Hill EW, Jobling MA, Bradley DG: Y-chromosome variation and Irish origins. *Nature* 2000; **404**: 351–352.
- 18 Gill P, Brenner C, Brinkmann B *et al*: DNA Commission of the international society of forensic genetics: recommendations on forensic analysis using Y-chromosome STRs. *Forensic Sci Int* 2001; **124**: 5–10.
- 19 Rodig H, Grum M, Grimmecke HD: Population study and evaluation of 20 Y-chromosome STR loci in Germans. *Int J Legal Med*, in press.
- 20 Kayser M, Caglia A, Corach D *et al*: Evaluation of Y-chromosomal STRs: a multicenter study. *Int J Legal Med* 1997; **110**: 125–133, 141–129.
- 21 Elmoznino M, Prinz M: Y-STR haplotype reference database [http://www.ystr.org/index\\_gr.html](http://www.ystr.org/index_gr.html), 2004.
- 22 Immel UD, Kleiber M, Klintschar M: Y chromosome polymorphisms and haplotypes in South Saxony-Anhalt (Germany). *For Sci Int* 2005 **155**: 211–215.
- 23 Excoffier L, Smouse PE, Quattro JM: Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 1992; **131**: 479–491.
- 24 Excoffier L, Smouse PE: Using allele frequencies and geographic subdivision to reconstruct gene trees within a species: molecular variance parsimony. *Genetics* 1994; **136**: 343–359.
- 25 Schneider S, Rosseli D, Excoffier LA: *A software for population genetics analysis (Ver. 2.000)*, genetics and biometry laboratory. Switzerland: University of Geneva, 2000.
- 26 Bartlett R: *The making of Europe*. London: Penguin, 1994.
- 27 Kohlheim R, Kohlheim V: *Familiennamen*. Mannheim: Duden, 2005.
- 28 Kunze K: *dtv-Atlas Namenkunde*. München: Deutscher Taschenbuch Verlag, 2004.