Forensic Science International: Genetics 7 (2013) 529–536

Contents lists available at SciVerse ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

High resolution mapping of Y haplogroup G in Tyrol (Austria)

Burkhard Berger [a,1], Harald Niederstätter [a,1], Daniel Erhart [a], Christoph Gassner [b,c], Harald Schennach [b], Walther Parson [a,d,*]

[a] *Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria*
[b] *Central Institute for Blood Transfusion & Immunological Department, Innsbruck, Austria*
[c] *Blood Transfusion Service Zürich, SRC, Schlieren, Switzerland*
[d] *Penn State Eberly College of Science, University Park, PA, USA*

ARTICLE INFO

Article history:
Received 27 February 2013
Received in revised form 8 May 2013
Accepted 30 May 2013

Keywords:
Human population genetics
G-L497
Y-SNP
Y-STR

ABSTRACT

The distribution of Y-chromosomal haplogroup G2a (G-P15) in present-day paternal lineages in Tyrol (Austria) was analyzed by applying a high-density regional sampling scheme that also covered remote mountain areas. There is evidence from ancient genetic data for a high frequency of Y-chromosomal haplogroup G in prehistoric populations of Central Europe, whilst nowadays levels well below 10% are routinely observed. A population sample comprising ~3700 specimens was analyzed for Y-chromosomal variation by genotyping Y-SNPs and Y-STRs. The set of binary markers included nine SNPs specific for sub-lineages of haplogroup G. The frequency of haplogroup G in 2379 unrelated men born in Tyrol amounted to 11.3%. Nearly all of these Y chromosomes belonged to haplogroup G2a. The main sub-haplogroup within G2a was defined by the SNP L497 (G2a3b1c) and reached a population frequency of 8.6%. Although this average level is higher than reported for other countries the geographical distribution of haplogroup G-L497 showed a differentiated pattern with a clustered distribution within some alpine valleys, where maxima above 40% were found. Both, the estimation of coalescent times and a principle coordinates analysis based on R_{ST} values derived from Y-STR haplotypes from different sub-regions of Tyrol revealed evidence for an old settlement history associated with Y chromosomes belonging to haplogroup G in the Tyrolean Alps.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

The analysis of Y chromosome variation is a well established tool for investigating human history on the basis of present-day genetic diversity. This approach performs well because global Y-chromosomal diversity is strongly geographically differentiated [1,2]. This also applies for different European populations, a fact known for more than one decade [3,4]. Regions can be characterized by their typical composition of Y-chromosomal haplogroups and the resulting spatial pattern is thought to arise from population expansions, migration events and the effect of genetic drift [5]. On more local scales, regional differences are also influenced by cultural practices such as patrilocality [6]. Studying ancient (a)DNA, e.g. obtained from archaeological bone samples, provides an alternative approach. This might become increasingly relevant in particular as the current progress in aDNA analysis makes direct comparisons with genetic data from archaeological remains feasible. There is strong evidence from ancient genetic data for a high frequency of Y-chromosomal haplogroup G in Neolithic populations of Central Europe [7–12], a haplogroup associated with the spread of agriculture on the European continent [4]. Combining all Neolithic DNA evidence revealed that this haplogroup was found in four out of six archaeological sites situated in Spain, France, Italy and Germany. The overall frequency of haplogroup G amounted to more than 70% (27 of 37 successfully typed Neolithic remains). The Neolithic expansion was a major event in the European settlement history and its role in shaping the present day genetic landscape of Europe is still the subject of discussion [13–18]. However, it is possible that bearers of aDNA sequences have no living descendants, and therefore limited relevance to today's genetic diversity. This brings us back to the importance of a detailed analysis of present-day genetic diversity. Nowadays, haplogroup G is most common in the Caucasus with maxima exceeding 70% and rapidly decreasing frequencies further eastward levelling off between 5 and 15% in both the Near/Middle East and in southern parts of Europe [19]. As

* Corresponding author at: Institute of Legal Medicine, Innsbruck Medical University, Innsbruck, Austria. Tel.: +43 512 9003 70651; fax: +43 512 9003 73640.
E-mail address: walther.parson@i-med.ac.at (W. Parson).

[1] These authors contributed equally to this work.

1872-4973/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2013.05.013

the frequency of haplogroup G and its sub-branches is generally low for the rest of Europe, a detailed picture of its distribution pattern demands a high sample number combined with a defined geographical resolution. The vast majority of European Y chromosomes associated to G belonged to G2 [20], more precisely to the G-L497 sub-haplogroup as recently demonstrated by Rootsi and colleagues [19]. The authors speculate that this lineage could potentially be associated with the Linear Pottery culture, a major archaeological horizon of the European Neolithic.

In this study we set out to test for levels of haplogroup G in present-day paternal lineages in Tyrol (Austria) by applying a high-density sampling regime to find local differentiation of Y chromosomes, even at a micro-geographic scale. Tyrol is a federal state in the western part of Austria. The state is split into two parts, a larger one called North Tyrol and smaller one commonly referred to as East Tyrol. North Tyrol borders on Germany (Bavaria) in the north, and Italy (South Tyrol, Trentino-Alto Adige) as well as Switzerland in the south (Fig. 1). The state's territory amounts 12,648 km^2 and is entirely located in the Eastern Alps. Due to Tyrol's alpine topology, its population of about 710,000 is largely restricted to the mountain-valleys. The landscape is characterized by a complex network of valleys of different size and direction enclosed between mountain ranges creating a fragmented settlement area with the Inn Valley as the densest populated region. Entering Austria in the very south-western part of Tyrol, the river Inn first runs eastwards through wide areas of Tyrol and its capital, Innsbruck, which is the largest urban settlement.

Concerning the distribution of haplogroup G the Tyrolean Alps are of special interest because they are situated in the central part of Europe building a barrier between northern and southern parts of Europe. Since the Last Glacial Maximum, valleys and mountain-passes were the major routes of migration. Different valleys can have highly unequal levels of isolation, one valley being a dead end with an ancient population history and the other one forming a transit route with a proposed higher level of admixture. In the case of Tyrol archaeological and historical records demonstrate this fact. The most prominent example is the "Tyrolean Iceman". The 5300 years old mummy was found in 1991 on the Similaun glacier in the Ötztal Alps near the border between Italy and Tyrol (Austria). The discovery of the Tyrolean Iceman proved that Neolithic men had already crossed the highest Alpine passes. Y-chromosomal single nucleotide polymorphism (Y-SNP) analyses assigned the Iceman to a sub-branch of G2a, which is defined by the SNP L91 [12]. This particular sub-haplogroup is very rare in present-day populations of Europe (<1%). However, the affiliation of the Iceman's Y chromosome to G proves that this Y-chromosomal haplogroup was present in this region in prehistoric times.

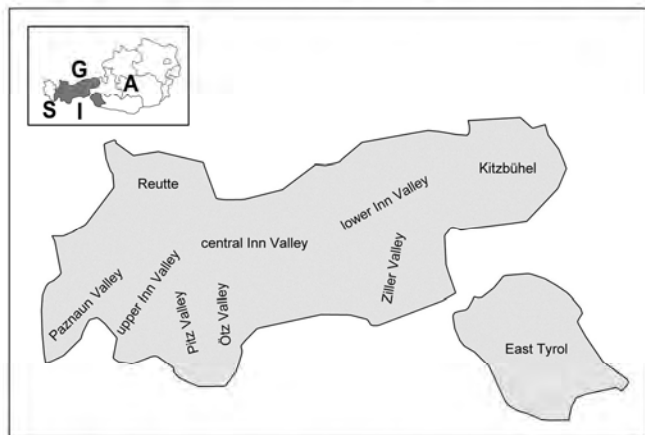

Fig. 1. Map of the Austrian federal state Tyrol. Selected area names are given within the map. A: Austria, G: Germany, I: Italy; S: Switzerland.

2. Materials and methods

2.1. Ethics statement, biological samples and data management

This study was reviewed and approved by the ethics commission of the Innsbruck Medical University (study classification number UN2598, session number 241/4.5).

Blood samples (~5 ml, $n = 3713$) from healthy West Eurasian men living in Tyrol (Austria) were obtained in the course of blood collection campaigns organized by the "Central Institute for Blood Transfusion and Immunological Department Innsbruck". All voluntary blood donors received detailed information about the study and provided both written consent as well as personal and genealogical data (full names, birth dates and the family's places of residence at the time of birth of the donor, his father and paternal grandfather). The personal data and all associated genotyping results were stored and managed by means of a dedicated database developed and held in-house.

Part of the genotyping data have been previously published [21,22] and were deposited in the Y Chromosome Haplotype Reference Database (YHRD; http://www.yhrd.org; [23,24]) under the accession numbers YA003716 (East Tyrol, Austria [Tyrolean]) and YA003715 (Reutte, Austria [Tyrolean]). In addition, internal quality assurance as well as regular participation in GEDNAP (German DNA Profiling) proficiency tests served to meet the standards of the International Society for Forensic Genetics for forensic genetic data [25]. This paper was written in accordance to the guidelines for publication of population data requested by the journal [26].

2.2. DNA extraction

Total genomic (g)DNA was extracted in the 96-well format from 200 μl blood-aliquots using the nexttec genomic DNA isolation kit for blood (nexttec Biotechnologie, Leverkusen, Germany) according to the manufacturer's instructions. The DNA samples were stored frozen until further use.

2.3. Y-SNP typing

A 19-plex polymerase chain reaction (PCR) amplification and a 19-plex single nucleotide primer extension (SNPE) assay (SNaPshot multiplex kit, Life Technologies, Paisley, UK) were used for the genotyping of the following phylogenetically informative SNPs on the male-specific region of the Y chromosome (MSY [27]): M9 (rs3900), M17 (rs3908), M45 (rs2032631), M78, M89 (rs2032652), M96 (rs9306841), M170 (rs2032597), M173 (P241, Page29, rs2032624), M201 (rs2032636), M223, M253 (rs9341296), M269 (rs9786153), M304 (Page16, rs13447352), M343 (rs9786184), P15, P37, SRY_{10831} (Page65, SRY_{1532}, rs2534636), U106 (M405, S21, rs16981293), and U152 (S28, rs1236440). For detailed protocols see [21].

2.3.1. KASPar Y-SNP genotyping

All samples featuring with 19-plex PCR-SNPE either M201 (haplogroup G, $n = 3$ [1]) or P15 (haplogroup G2a, $n = 413$ [28]) as the most derived Y-SNP were further genotyped for the markers L32 (S148, U8, rs7892988 [20]) and L497 (S317, rs35141399 [19]) using homogeneous competitive allele-specific PCR followed by endpoint-fluorescence detection (KASPar, KBioscience, Hoddesdon, UK) [29].

KASPar assays (8 μl) comprised 1× Kasp ver. 4.0 SNP genotyping reaction mix (low ROX), 0.11 μl Kasp on Demand assay mix (both KBioscience), and ~2 ng gDNA.

Each of the two KASPar assays (L32 and L497) constituted a pair of allele-specific primers carrying distinct non-homologous

5′ sequence tails, one common opposite-strand primer (Table S1), two 5′ fluorescent labelled (FAM or VIC) reporter-primers exhibiting the same sequence as the unique tags attached to the allele-specific primers, and two oligonucleotides with fluorescence quenchers at their 3′ ends. The latter were designed to bind non-incorporated reporter-primers.

All amplifications were conducted on an Applied Biosystems 7500 Fast real-time PCR system (Life Technologies) and comprised initial denaturation at 94 °C for 15 min, followed by 10 cycles of 94 °C for 20 s and 61 °C for 1 min (−0.6 °C/cycle) and 26 cycles of 94 °C for 10 s and 55 °C for 1 min. For endpoint fluorescence based allelic discrimination, the Applied Biosystems Sequence Detection Software (version 1.4, Life Technologies) was used.

2.3.2. Chain termination sequencing

All G-P15 specimens featuring the derived base state at L32 and the ancestral allele at L497 were further typed for the Y-SNPs M406 [30] and L645 (International Society of Genetic Genealogy (2012), Y-DNA haplogroup Tree 2013, http://www.isogg.org/tree) by means of direct chain termination sequencing of PCR products.

All G-P15 samples exhibiting the ancestral allelic state at both L32 and L497 were further analyzed for the Y-SNPs L293 (ISOGG 2013), P16 [28], M286 (rs13447379 [31]), and L91 (S285 [12]) using direct Sanger-type sequencing of PCR products.

The phylogenetic tree defined by this haplogroup G specific Y-SNP panel is shown in Fig. 2.

Finally, part of the KASPar allele calls (L32: $n = 49$, L497: $n = 49$) were confirmed by sequencing analysis.

2.3.2.1. PCR and cycle sequencing. All singleplex PCR cocktails (20 μl) comprised of 1× buffer II (Life Technologies), 5 μg non-acetylated BSA (Sigma–Aldrich, St. Louis, MO, USA), 1.5 mM $MgCl_2$, 200 μM each dNTP, 250 nM forward primer, 250 nM reverse primer (Microsynth, Balgach, CH, Table S1), 2.5 U AmpliTaq Gold DNA polymerase (Life Technologies), and ∼2 ng gDNA.

Amplifications were performed in 96-well polypropylene PCR plates on an Applied Biosystems 9700 thermal cycler (Life Technologies). The thermal cycler protocol comprised initial denaturation at 95 °C for 10 min, followed by 38 cycles of 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 45 s. The final extension step was extended by 10 min. For two target sequences the three temperature repeat-stage was replaced by a two-step protocol using 38 cycles of denaturation at 95 °C for 15 s and annealing/extension at 65 °C (M286) or 69 °C (L293) for 1 min.

Enzymatic post-PCR treatment, cycle sequencing using the amplification primers (Table S1) as sequencing primers (assay concentration: 160 nM), clean-up of the cycle sequencing products by size exclusion chromatography, and laser-induced fluorescence detection capillary electrophoretic separation were performed according to [32]. The sequencing traces were analyzed with the Sequencher software (version 5, Gene Codes, Ann Arbor, MI, USA).

2.4. Y-STR typing

Seventeen-locus Y-chromosomal short tandem repeat (Y-STR) haplotypes were determined with the AmpFlSTR Yfiler PCR amplification kit (Life Technologies) as described in [33].

Laser induced fluorescence detection capillary electrophoretic separation of PCR products was performed on an ABI 3100 Genetic Analyzer, using a 36 cm capillary array, POP-6 (all Life Technologies) as sieving matrix, and default conditions for sample loading and separation. Raw-data were analyzed with the GeneMarker HID computer program (version 1.70, SoftGenetics, State College, PA, USA).

The nomenclature for all STR markers followed the recommendations of the International Society for Forensic Genetics [34].

2.5. Data analysis

2.5.1. Sample filtering

The sample set comprising genotypes of all donors was filtered for possible relatives (kinship filter) and regarding to the geographic origin and known duration of residence in Tyrol (residence filter).

2.5.1.1. Kinship filter. Groups of samples featuring both identical Y-STR haplotypes (on basis of the 17 Y-STR markers included in the Yfiler kit) and identical surnames were considered to derive from close paternal relatives. In all such cases, only one specimen per group was retained in the sample set.

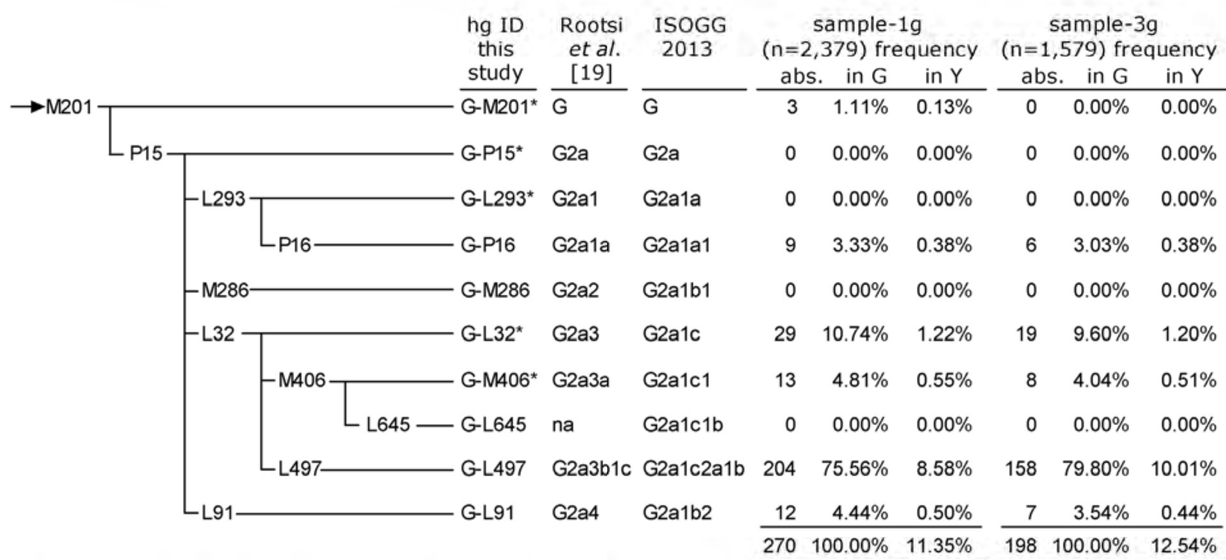

Fig. 2. Binary Y-SNP haplogroup tree and Y-chromosomal haplogroup frequencies. The phylogenetic tree depicts the relationships among the binary Y-chromosomal haplogroups defined by the 10 haplogroup G specific SNPs analyzed in this study. Haplogroup designations are according to the Y-Chromosome Consortium (YCC) nomenclature principles [47] and were taken from [19] and the ISOGG 2013 tree (International Society of Genetic Genealogy 2013, http://www.isogg.org/tree/index.html). Absolute and relative frequencies of haplogroup G (sub)lineages found in Y chromosomes comprised in sample-1g and sample-3g are listed right-hand of the tree.

2.5.1.2. Residence filter. Unrelated men who declared to be born in Tyrol formed the one-generation of residence sample set (hereafter referred to as "sample-1g"). Donors who declared Tyrolean residence dating back to their paternal grandfather were summarized as three generations of residence sample set ("sample-3g") and, therefore, constituted a sub-sample of sample-1g.

2.5.2. Coalescent time estimates

Coalescent times (T_d) of sub-haplogroup G-L497 in specific regions of Tyrol were estimated using the method described by [19,35] using an evolutionary effective mutation rate of 6.9×10^{-4} and 25 years as average generation time. The calculations were based on the variation of Y-STR loci and, in particular, on the average squared difference in the number of repeats between all current chromosomes of a sample and the founder haplotype, which was defined as the median of the current haplotypes as proposed by [36]. For this purpose for each STR locus the median value of repeat scores was computed in order to form a median haplotype. Eight STR markers of our STR set (DYS19, DYS389I, DYS389b, DYS390, DYS391, DYS392, DYS393, and DYS439) overlapped with those used by [19] and were, therefore, selected for coalescent time estimates, to achieve the most reliable comparison possible with the data computed by these authors.

2.5.3. Database queries

For database queries using the YHRD (release 41) the most frequent allele of each STR marker was combined in order to mimic a hypothetic founder haplotype for haplogroup G-L497, which exactly corresponded to the median haplotype described above. The median haplotype of the so-called minimal haplotype (minHT) consisting of the loci DYS19, DYS398I, DYS389II, DYS390, DYS391, DYS392, DYS393, and DYS385a/b was computed for all G-L497 Y chromosomes. In order to include haplotypes closely related to the median minHT, all those haplotypes with a deviation of one repeat step up or down in one marker (neighbour haplotypes) were included. The resulting cluster consisting of the median minHT and all corresponding neighbour haplotypes is hereinafter referred to as "median HT cluster" (Table S2).

2.5.4. Population genetics analyses of Y-STR haplotypes

For all calculations the inferred Y-STR repeat unit counts were converted into repeat block lengths, DYS389b was used instead of DYS389II and both DYS385 alleles were removed from the Y-STR profiles. Any other duplicated Y-STRs were treated as missing data.

The Arlequin software (version 3.5.1.2) [37] was used for the computation of haplotype pairwise R_{ST} distances. To assess the statistical significance ($P < 0.05$) of R_{ST} distances, the number of permutations was set to 10,000, and the method published by [38] was employed to correct for multiple comparisons (Table S3).

GENALEX 6.4 [39] was used for principal coordinates analysis (PCA).

2.5.5. Spatial frequency maps

Spatial frequency maps were obtained by applying frequency data to the MapViewer software (Version 7.3, Golden Software, Inc., Golden, CO, USA) using kriging interpolation. For calculating frequency data of G-L497 for the individual Tyrolean municipalities a threshold of 10 samples per municipality was arbitrarily set for sample-1g and sample-3g. Thus, only subsets of the total G-L497 sample were applied to the frequency maps (for exact numbers see legends to Figs. 3 and S2).

3. Results and discussion

More than 99% ($n = 3701$) of the total 3713 samples obtained from voluntary donors during blood collection campaigns covering

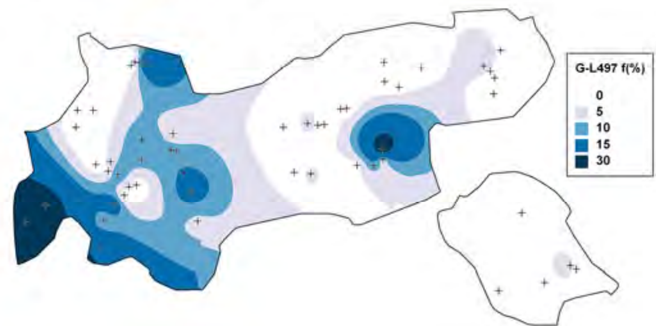

Fig. 3. Spatial frequency distribution of haplogroup G-L497 in Tyrol derived from the sample-1g data basis. The map shows the geographical distribution of 161 G-L497 Y-chromosomes found in a total of 1805 donors born in 54 different Tyrolean municipalities (indicated by "+") with a sample size > 10.

all districts of Tyrol, could be successfully analyzed for both Y-SNPs and Y-STRs. This sample set was filtered by applying both a residence as well as a kinship filter. In a first step only donors that were born in Tyrol were considered, showing that ~92% ($n = 3404$) of the successfully typed sample set originated from donors born in Tyrol.

A filter for close paternal relatives appeared to be essential for the following two reasons. (1) The sampling regime included particularly small municipalities from relatively secluded rural locations, which increased the probability of coincidental sampling of relatives. (2) A sampling bias caused by family traditions to attend blood sampling campaigns, resulting in an overrepresentation of the associated paternal lineages, could not be ruled out. However, the strength of the kinship filter needed to be adjusted carefully to preserve the demographic characteristics of the sample set with respect to the level of distant relatives, which can be found to varying extent in all populations [40]. In small villages and rural regions we expected increased patrilocality and, therefore, elevated levels of individuals who are cryptically related. Such cousins (e.g. on the order of 2nd to 9th grade) often show indistinguishable Y-chromosomal genotypes, even on the basis of Y-STRs having a high differentiating capacity [41]. A too stringent kinship filter obscures this characteristic of the population and distorts the results towards a higher diversity of paternal lineages. On the other hand, an over-reduction of genetic variability, e.g. by filtering new mutations within lineages of relatives should be minimized as well as the exclusion of DNA samples that are identical by state. To consider these partly opposing requirements, a combination of cultural and genetic markers for relatives was used, particularly the surnames of the donors and the corresponding Y-STR-haplogroups. This combination seems reasonable, because patrilineal surnames and Y chromosomes are passed in a similar manner and there is a strong relationship between the Y-STR haplogroup and the surname [42–45]. Surnames as one part of the kinship filter will minimize the exclusion of data that are identical by state. On the other hand, an appropriate set of Y-STR markers prevents an over-reduction of the data set, as their rapid mutation rates can lead to high haplotype diversity even among Y chromosomes that share a relatively recent common ancestor [41]. Calculations based on mutation rate estimates of the 17 Y-STR markers included in the Yfiler kit [46] showed that identical Yfiler genotypes capture more than 90% of father–son and brother pairs but the overall filter rate for relatives up to 20 generations amounted only ~85%. Therefore, Yfiler genotypes are very efficient to detect close relatives but provide adequate scope for cryptic distant relatives.

Groups of samples featuring both identical Y-STR haplotypes (on basis of Yfiler markers) and surnames were considered to derive from close paternal relatives. In all such cases, only one

specimen per group was retained in the sample set. Finally, the population sample comprised 2379 specimens from, according to this filter, paternally unrelated men who declared to be born in Tyrol (sample-1g). A portion of 66% (1579 specimens) was donated by probands who declared Tyrolean residence dating back to their paternal grandfather (sample-3g).

Sample-1g consisted of representatives from nearly all parts of the country, which was supported by the fact that 225 out of 279 (81%) Tyrolean municipalities were covered. This also applied to sample-3g. In this case the donors' paternal grandfathers came from even slightly more Tyrolean municipalities ($n = 226$).

The two sample sets allowed the analysis of the Y-chromosomal landscape present in the mid 20th and late 19th century, as the median year of birth of the donors and their paternal grandfathers was 1962 and 1898, respectively.

Y-chromosomal SNP typing resulted in 270 Y chromosomes belonging to haplogroup G (G-M201) (Table S4) for sample-1g representing a fraction of 11.3%, a value that was roughly two to three times higher than those reported for most of the European countries [19]. For Switzerland, Sicily and Italy similar levels were reported, only Corsica and Sardinia showed higher frequencies [19]. In the Tyrolean districts Lienz (East Tyrol) and Reutte G-P15 frequencies of 7.4% [21] and 8.5% [22] were found, respectively.

By typing a set of nine SNP loci specific for different G sub-haplogroups, the 270 Y chromosomes belonging to haplogroup G were assigned to four of the nine theoretically distinguishable haplo/paragroups (see Fig. 2). In three samples only M201 showed a derived allele. Therefore, it was not possible to assign these Y chromosomes to specific sub-haplogroups within G. Hence, they had to be called G-M201* within the framework of our SNP set. Considering the comprehensive SNP data presented by [19] it seems justified to speculate that these three samples could belong to one or more sub-haplogroup(s) found in Europe by these authors.

The sub-haplogroup G-L91, which was assigned to the Y chromosome of the Tyrolean Iceman [12], was found in 12 unrelated men born in Tyrol representing a mere 0.5% of the male Tyrolean population sample. Therefore, this sub-haplogroup constitutes in terms of quantity no significant part of the Y-chromosomal make-up of present-day Tyrol. With the exception of G-L497, the same applied to the remaining sub-haplogroups, which showed population frequencies of less than 2% in sample-1g and sample-3g. In contrast, G-L497 ($n = 204$) comprised the major proportion (~75%) of haplogroup G Y chromosomes (Fig. 2). A similar composition was found for sample-3g, indicating that no profound changes within the population with respect to these haplogroups occurred since the end of the 19th century (Fig. 2).

The G-L497 frequencies of 8.6% and 10% found for sample-1g and sample-3g, respectively, were higher than in all populations from Europe, the Near/Middle East and the Caucasus as listed by [19]. This supports the assumption of Rootsi and co-workers that this haplogroup originated from Central Europe and that the occurrence of this lineage in Italy could be caused by migratory flows from northern parts of the continent [19]. This, in turn, pointed out that the Alps, and in particular the Tyrolean Alps might have played a significant role as source or at least as transit route of this lineage.

The spatial distribution of G-L497 Y chromosomes within Tyrol was studied by calculating the frequencies in different Tyrolean municipalities. A minimum required sample size of 10 was arbitrarily defined to allow for reliable frequency estimates on the level of individual municipalities. For 54 municipalities sample sizes between 98 and ten samples each were obtained and were, therefore, used for frequency calculations.

Haplogroup G-L497 was anything but evenly distributed. As shown in Fig. S1 in one third of the municipalities ($n = 18$) no G-L497 Y chromosomes were found, but on the other hand frequencies reached levels exceeding 40% in other regions of Tyrol. The spatial frequency map of haplogroup G-L497 indicated a distinct distribution pattern with areas of high G-L497 levels in certain alpine valleys, particularly in the Paznaun Valley, the Upper Inn Valley, the Ötz Valley and the Ziller Valley (see Figs. 3 and S2; for locations see Fig. 1).

The deduction of historic events from the present-day patterns of genetic diversity is limited by the possibility that the ancient picture became blurred by more recent events. Isolation, founder effect and genetic drift, arising through stochastic variation in the number of offspring are factors that are known to cause marked population differentiation, particularly for the uniparentally inherited segments of the Y chromosome. The effects are stronger for this part of the genome, because on population levels they are fewer in number than autosomal segments. Due to the sex-specific pattern of inheritance of the Y chromosome, the effect of the above mentioned factors could be enhanced by social factors such as patrilocality. Therefore, the question arises whether or not the observed patterning of modern G-L497 Y chromosomes can be utilized for drawing conclusions on the early settlement. By connecting the recent genetic data with the personal data of the donors' paternal grandfathers a picture of the G-L497 distribution at the end of the 19th century could be deduced. The pattern obtained (Fig. S2) resembled the present-day results (Fig. 3) and makes it thus unlikely that the high frequencies of G-L497 in specific regions of Tyrol were the result of very recent events, like effects of industrialization, modern transport infrastructure or tourism. Yet, for drawing conclusions regarding periods longer ago than approximately one century the analysis of the spatial distribution of extant G-L497 Y chromosomes is not applicable. For this purpose Y-STR variability was used in an attempt to distilling information regarding haplogroup relationships among different geographic (sub)regions and to estimate the age of the haplogroups within these regions.

For all 270 specimens belonging to haplogroup G full 17-locus Y-STR haplotypes were obtained (Table S4). Ten sub-regions of Tyrol were defined taking into account the geographic distinctiveness of the areas as well as the number of available samples (see marks on the map shown in Fig. 1). Pairwise genetic distances (R_{ST}) were computed on the basis of the Y-STR haplotypes obtained for the ten sub-regions. To display and quantify the major patterning in the Y-STR data, the relationships among the pairwise R_{ST} values were plotted by means of PCA (Fig. 4). About 97% of the total haplotypic variation could be attributed to the first (~82%) and second (~15%) principal coordinates. More than half of all sub-regions constituted a cluster containing all parts of the Inn Valley as well as the districts Reutte, Kitzbühel and East Tyrol, situated in the most northern, eastern and southern part of the county, respectively. It is worth noting, that the Y-STR haplotypes from the valleys showing the highest G-L497 levels were clearly distinct from the main cluster and from each other as well. The clear phylogeographic differentiation among at least five different regions takes place at a microgeographic scale. For instance, the Paznaun Valley, the upper Inn Valley, the Pitz Valley and the Ötz Valley, all featuring G-L497 levels above average, are located within a radius of ~60 km in the western part of Tyrol. This patterning seen with PCA obviously pointed against a recent settlement of Tyrol by men carrying a G-L497 Y chromosome. In contrast, it makes the conclusion plausible, that these Y chromosomes might have an independent, deep-rooting history. This postulation was further supported by age estimates for haplogroup G-L497. With an estimate of the mutation rate of Y-STRs, it is possible to estimate how much time elapsed since the most recent common ancestor of a specific haplogroup lived [35]. Using Y-STR data from different regions of Tyrol, the coalescent expansion

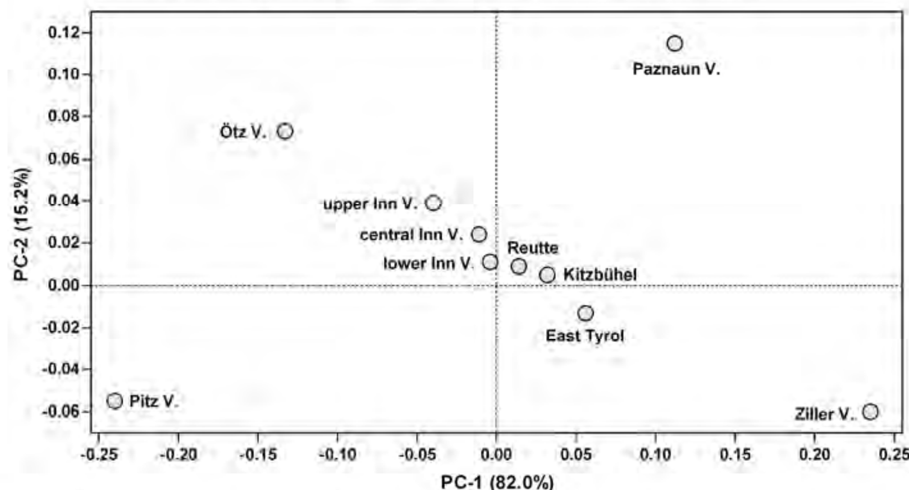

Fig. 4. Principal coordinates analysis. Plot showing the first and second principal coordinates determined by PCA using pairwise Slatkin linearized R_{ST} distances, as obtained for the Y-STR haplotypes comprised in sample-1 g subdivided in a set of 10 regional areas of Tyrol as shown in the map (Fig. 1).

times for the L497 affiliated Y chromosomes were estimated as listed in Table 1. The coalescent time estimate for the whole Tyrolean sample set (on the basis of sample-1g) amounted 7500 years with a standard error of ±1300 years. This perfectly matched the coalescent time-estimate of 7300 ± 1600 years reported for whole Europe [19], even though our results based on 8 STRs cannot be exactly compared to those of Rootsi and colleagues, who used 10 STR markers for their coalescent time estimates. Generally, such calculations based on microsatellite variance have to be interpreted with caution as it is known that convergence time measures obtained from other sources, e.g. sequence data, may provide deviating results. Despite these considerations, it has been shown that microsatellite variance can be used to identify realistic time frames for the formation of Y chromosomal lineages [35,36]. The coalescent time-estimates obtained for the different Tyrolean sub-regions varied between 3800 and 13,900 years (Table 1) and, hence, suggest an old settlement history of G-L497 Y chromosome carriers in the different Tyrolean valleys.

Our results clearly indicate that the Tyrolean Alps are of significance for the geographical extension of G-L497 as shown by the high abundance, microsatellite diversity and estimated age of this particular haplogroup. However, little is known about the detailed distribution of this haplogroup in Europe and in particular in the neighbouring areas of Tyrol. The distribution of G-L497 presented by [19] gives a first overview. However, until now SNP data with a high sub-haplogroup resolution are still missing for

Table 1
Coalescent time (T_d) estimations of haplogroup G-L497 Y chromosomes from Tyrol and different sub-regions. T_d and the standard error (SE) are given in thousand years [kY]. For comparison the T_d estimate for a European population sample reported by [19] is included.

Population	T_d [kY]	SE	n
Tyrol (total)	7.3	1.3	204
Tyrol sub-regions			
Pitz Valley	3.8	1.6	6
Reutte	5.4	1.2	10
Ötz Valley	5.6	2.1	38
Central Inn Valley	5.7	1.2	19
Upper Inn Valley	6.3	1.7	45
Kitzbühel	7.2	2.1	10
Lower Inn Valley	8.2	1.5	20
Ziller Valley	9.3	3.3	18
Paznaun Valley	10.8	4.9	26
East Tyrol	13.9	3.3	12
Europe [19]	7.5	1.6	63

many parts of Europe. To make use of a broader data foundation we used Y-STR haplotypes considered to be representative for G-L497 for a database search in the YHRD. By combining the most frequent allele of each STR marker within the G-L497 data set we aimed to approach the (hypothetical) founder haplotype for this haplogroup. In that sense, the resulting haplotype was analogously determined as a modal haplotype and precisely corresponded to the median haplotype calculated for coalescent time estimations. This is not unexpected for a haplogroup that has existed for a sufficiently long time [36]. On the level of the minHT this G-L497 specific haplotype was determined as 15–12–29–22–10–11–14–14/14. With reference to the obvious consistency to the median haplotype it was hereafter referred to as "median minHT". To allow for a deviation to the extent of a single mutation all one-step neighbours of this haplotype were included in the search (median HT cluster). This group of haplotypes proved to be highly abundant in G-L497 as it was established in 43% of all 204 G-L497 Y chromosomes of our study. In the remaining 2175 samples it was found only three times (0.11%). The YHRD search comprised 173 European populations kept in the database with sample sizes > 50. The median HT cluster was found 260 times in 98 different populations comprising of 34,386 samples (see Table S2). This corresponded to an average frequency of 0.8% for whole Europe. The spatial distribution of the median HT cluster within Europe is shown in Fig. 5. By far the highest frequency was found for the Mediterranean island Ibiza (18%; n = 96), followed by Verona (4%; n = 153) and Tyrol (4%; n = 230). The Tyrolean sample in the YHRD is independent from the data presented in this study, but the level of the median HT cluster in the 2379 specimens of sample-1g is nearly the same (3.8%). The distribution map shows a differentiated pattern with increased frequencies in the western Mediterranean Balearic islands, Sicily, the Carpathian Mountains, and the Alps and at their southern edge in parts of North Italy. The fact that all regions with a high frequency of the G-L497 specific median HT cluster exhibit a certain level of isolation, by being either an island or mountainous, suggests that these areas could constitute refuge areas for formerly abundant and widespread G-L497 Y chromosomes. A similar consideration also applies to another sub-haplogroup of G, namely G-L91, which is rare in mainland Europe but quite frequent in the islands of Corsica and Sardinia [12]. However, this remains a speculation unless new facts corroborate our idea.

The high-resolution regional sampling performed in this study on Tyrolean Y chromosomes revealed a differentiated spatial distribution of haplogroup G-L497 with unexpectedly high levels

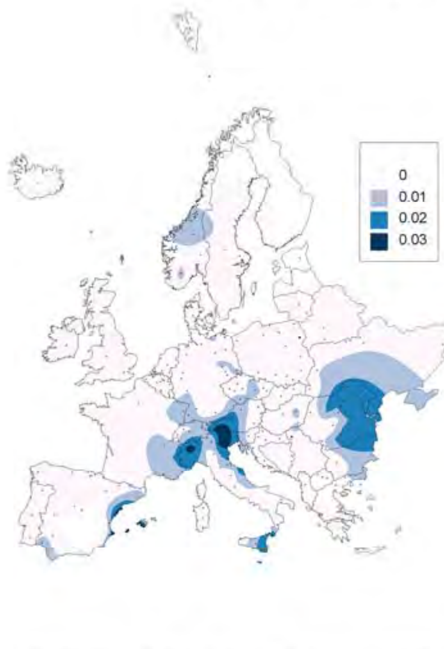

Fig. 5. Spatial frequency distribution of the median HT cluster specific for haplogroup G-L497 in Europe. The frequency data were obtained by a search in the YHRD (release 41) and comprise 260 samples belonging to the median HT cluster found in 34,386 samples from 173 European population samples (sample size > 50). The sample locations are indicated by "+".

in some mountain valleys. In general, this sampling strategy enabled us to identify details of Y-chromosomal variation even in haplogroups that are not common on a larger scale. The detailed results obtained from the Tyrolean sample help to get a more complete picture of haplogroup G in Europe and support some considerations regarding its settlement history as discussed in detail above. However, the indications that high levels of haplogroup G-L497 are characteristic for residents of the Alps in general has to be verified by additional Y chromosomal analyses of men living in neighbouring countries, e.g. North Italy and Switzerland.

Conflicts of interest

The authors declare that they do not have any known conflict of interest.

Acknowledgements

The authors would like to thank all voluntary blood donors. The staff of the "Central Institute for Blood Transfusion and Immunological Department" (Landeskrankenhaus Innsbruck) is acknowledged for organizing the blood collection campaigns. We are obliged to Stefan Troger (Institute of Legal Medicine, Innsbruck Medical University) for programming and maintaining the in-house database. Hannelore Volderauer (Institute of Legal Medicine, Innsbruck Medical University) is greatly acknowledged for sample management and administration. This work was financially supported by the Grants "ALPKULTUR – cultural historical documentation of toponyms in the alpine area" and "High-Resolution Analysis of Y chromosome Variability in Selected Regions of the Tyrolean Alps (Austria): Influences of Topology and Demographic History on the Genetic Differentiation" within the "Alpine Research Programme" of the Austrian Academy of Sciences. This work leading to these results has also received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 285487 (EUROFORGEN-NoE). D.E. received a scholarship from the Leopold Franzens University of Innsbruck.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2013.05.013.

References

[1] P.A. Underhill, G. Passarino, A.A. Lin, P. Shen, M. Mirazón Lahr, R.A. Foley, P.J. Oefner, L.L. Cavalli-Sforza, The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations, Ann. Hum. Genet. 65 (2001) 43–62.

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

[3] Z.H. Rosser, T. Zerjal, M.E. Hurles, et al., Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language, Am. J. Hum. Genet. 67 (2000) 1526–1543.

[4] O. Semino, G. Passarino, P.J. Oefner, et al., The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective, Science 290 (2000) 1155–1159.

[5] M.A. Jobling, The impact of recent events on human genetic diversity, Philos. Trans. R. Soc. Lond. B: Biol. Sci. 367 (2012) 793–799.

[6] S.J. Marks, H. Levy, C. Martinez-Cadenas, F. Montinaro, C. Capelli, Migration distance rather than migration rate explains genetic diversity in human patrilocal groups, Mol. Ecol. 21 (2012) 4958–4969.

[7] W. Haak, G. Brandt, H.N. de Jong, C. Meyer, R. Ganslmeier, V. Heyd, C. Hawkesworth, A.W. Pike, H. Meller, K.W. Alt, Ancient DNA, strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 18226–18231.

[8] W. Haak, O. Balanovsky, J.J. Sanchez, et al., Members of the Genographic Consortium, ancient DNA from European early neolithic farmers reveals their near eastern affinities, PLoS Biol. 8 (2010) e1000536.

[9] M. Lacan, C. Keyser, F.X. Ricaut, N. Brucato, J. Tarrús, A. Bosch, J. Guilaine, E. Crubézy, B. Ludes, Ancient DNA suggests the leading role played by men in the Neolithic dissemination, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 18255–18259.

[10] M. Lacan, C. Keyser, F.X. Ricaut, N. Brucato, F. Duranthon, J. Guilaine, E. Crubézy, B. Ludes, Ancient DNA reveals male diffusion through the Neolithic Mediterranean route, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 9788–9791.

[11] E.J. Lee, C. Makarewicz, R. Renneberg, et al., Emerging genetic patterns of the European Neolithic: perspectives from a late Neolithic Bell Beaker burial site in Germany, Am. J. Phys. Anthropol. 148 (2012) 571–579.

[12] A. Keller, A. Graefen, M. Ball, et al., New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing, Nat. Commun. 3 (2012) 698.

[13] M. Currat, L. Excoffier, The effect of the Neolithic expansion on European molecular diversity, Proc. Biol. Sci. 272 (2005) 679–688.

[14] G. Barbujani, L. Chikhi, Population genetics: DNAs from the European Neolithic, Heredity 97 (2006) 84–85.

[15] B. Bramanti, M.G. Thomas, W.W. Haak, et al., Genetic discontinuity between local hunter-gatherers and central Europe's first farmers, Science 326 (2009) 137–140.

[16] P. Balaresque, G.R. Bowden, S.M. Adams, et al., A predominantly neolithic origin for European paternal lineages, PLoS Biol. 8 (2010) e1000285.

[17] P. Soares, A. Achilli, O. Semino, W. Davies, V. Macaulay, H.J. Bandelt, A. Torroni, M.B. Richards, The archaeogenetics of Europe, Curr. Biol. 20 (2010) 174–183.

[18] R. Rasteiro, L. Chikhi, Female and male perspectives on the neolithic transition in Europe: clues from ancient and modern genetic data, PLoS ONE 8 (2013) e60944.

[19] S. Rootsi, N.M. Myres, A.A. Lin, et al., Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus, Eur. J. Hum. Genet. 20 (2012) 1275–1282.

[20] L.M. Sims, D. Garvey, J. Ballantyne, Improved resolution haplogroup G phylogeny in the Y chromosome, revealed by a set of newly characterized SNPs, PLoS ONE 4 (2009) e5792.

[21] H. Niederstätter, G. Rampl, D. Erhart, F. Pitterl, H. Oberacher, F. Neuhuber, I. Hausner, C. Gassner, H. Schennach, B. Berger, W. Parson, Pasture names with Romance and Slavic roots facilitate dissection of Y chromosome variation in an exclusively German-speaking alpine region, PLoS ONE 7 (2012) e41885.

[22] D. Erhart, B. Berger, H. Niederstätter, C. Gassner, H. Schennach, W. Parson, Frequency data for 17 Y-chromosomal STRs and 19 Y-chromosomal SNPs in the Tyrolean district of Reutte, Austria, Int. J. Legal Med. 126 (2012) 977–978.

[23] L. Roewer, M. Krawczak, S. Willuweit, et al., Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci. Int. 118 (2001) 106–113.

[24] S. Willuweit, L. Roewer, International forensic Y chromosome user group. Y chromosome haplotype reference database (YHRD): update, Forensic Sci. Int. Genet. 1 (2007) 83–87.

[25] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci. Int. 165 (2007) 238–243.

[26] A. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, P.M. Schneider, New guidelines for the publication of genetic population data, Forensic Sci. Int. Genet. (2013), http://dx.doi.org/10.1016/j.fsigen.2013.01.

[27] H. Skaletstky, T. Kuroda-Kawaguchi, P.J. Minx, et al., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes, Nature 423 (2003) 825–837.

[28] M.F. Hammer, A.J. Redd, E.T. Wood, et al., Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 6769–6774.

[29] E. Cuppen, Genotyping by allele-specific amplification (KASPar), CSH Protoc. (2007), http://dx.doi.org/10.1101/pdb.prot4841.

[30] R.J. King, S.S. Ozcan, T. Carter, et al., Differential Y-chromosome Anatolian influences on the Greek and Cretan Neolithic, Ann. Hum. Genet. 72 (2008) 205–214.

[31] C. Cinnioğlu, R. King, T. Kivisild, et al., Excavating Y-chromosome haplotype strata in Anatolia, Hum. Genet. 114 (2004) 127–148.

[32] A. Brandstätter, H. Niederstätter, M. Pavlic, P. Grubwieser, W. Parson, Generating population data for the EMPOP database – an overview of the control region as example, Forensic Sci. Int. 166 (2007) 164–175.

[33] B. Berger, A. Lindinger, H. Niederstätter, P. Grubwieser, W. Parson, Y-STR typing of an Austrian population sample using a 17-loci multiplex PCR assay, Int. J. Legal Med. 119 (2005) 241–246 (Erratum in: Int. J. Legal Med. 120 (2006) 255).

[34] L. Gusmão, J.M. Butler, A. Carracedo, P. Gill, M. Kayser, W.R. Mayr, N. Morling, M. Prinz, L. Roewer, C. Tyler-Smith, P.M. Schneider, International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis, Int. J. Legal Med. 120 (2006) 191–200.

[35] L.A. Zhivotovsky, P.A. Underhill, C. Cinnioğlu, et al., The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time, Am. J. Hum. Genet. 74 (2004) 50–61.

[36] S. Sengupta, L.A. Zhivotovsky, R. King, et al., Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists, Am. J. Hum. Genet. 78 (2006) 202–221.

[37] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, Mol. Ecol. Resour. 10 (2010) 564–567.

[38] Y. Benjamini, D. Yekutieli, The control of false discovery rate under dependency, Ann. Stat. 29 (2001) 1165–1188.

[39] R. Peakall, P.E. Smouse, GENALEX 6: genetic analysis in excel. Population genetic software for teaching and research, Mol. Ecol. Notes 6 (2006) 288–295.

[40] B.M. Henn, L. Hon, J.M. Macpherson, N. Eriksson, S. Saxonov, I. Pe'er, J.L. Mountain, Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples, PLoS ONE 7 (2012) e34267.

[41] K.N. Ballantyne, M. Goedbloed, R. Fang, et al., Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am. J. Hum. Genet. 87 (2010) 341–353.

[42] T.E. King, S.J. Ballereau, K.E. Schürer, M.A. Jobling, Genetic signatures of coancestry within surnames, Curr. Biol. 16 (2006) 348–384.

[43] G.R. Bowden, P. Balaresque, T.E. King, et al., Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England, Mol. Biol. Evol. 25 (2008) 301–309.

[44] T.E. King, M.A. Jobling, Founders, drift, and infidelity: the relationship between Y chromosome diversity and patrilineal surnames, Mol. Biol. Evol. 26 (2009) 1093–1102.

[45] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, Trends Genet. 25 (2009) 351–360.

[46] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, Forensic Sci. Int. Genet. 6 (2012) 208–218.

[47] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838.