Forensic Science International: Genetics 5 (2011) 95–99

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Micro-geographic distribution of Y-chromosomal variation in the central-western European region Brabant

Maarten H.D. Larmuseau [a,b,c,*], Nancy Vanderheyden [a], Manon Jacobs [a], Monique Coomans [a], Lucie Larno [a], Ronny Decorte [a,b]

[a] UZ Leuven, Department of Forensic Medicine, Laboratory of Forensic Genetics and Molecular Archaeology, Kapucijnenvoer 33, B-3000 Leuven, Belgium
[b] Katholieke Universiteit Leuven, Department of Human Genetics, Campus Gasthuisberg, Herestraat 49, B-3000 Leuven, Belgium
[c] Katholieke Universiteit Leuven, Laboratory of Animal Diversity and Systematics, Deberiotstraat 32, B-3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 25 June 2010
Received in revised form 12 August 2010
Accepted 25 August 2010

Keywords:
Genealogy
Low Countries
Population genetics
West-Europe
Y-chromosome

ABSTRACT

One of the future issues in the forensic application of the haploid Y-chromosome (Y-chr) is surveying the distribution of the Y-chr variation on a micro-geographical scale. Studies on such a scale require observing Y-chr variation on a high resolution, high sampling efforts and reliable genealogical data of all DNA-donors. In the current study we optimised this framework by surveying the micro-geographical distribution of the Y-chr variation in the central-western European region named Brabant. The Duchy of Brabant was a historical region in the Low Countries containing three contemporary Belgian provinces and one Dutch province (Noord-Brabant). 477 males from five *a priori* defined regions within Brabant were selected based on their genealogical ancestry (known pedigree at least before 1800). The Y-haplotypes were determined based on 37 Y-STR loci and the finest possible level of substructuring was defined according to the latest published Y-chr phylogenetic tree. In total, eight Y-haplogroups and 32 different subhaplogroups were observed, whereby 70% of all participants belonged to only four subhaplogroups: R1b1b2a1 (R-U106), R1b1b2a2* (R-P312*), R1b1b2a2g (R-U152) and I1* (I-M253*). Significant micro-geographical differentiation within Brabant was detected between the Dutch (Noord-Brabant) *vs.* the Flemish regions based on the differences in (sub)haplogroup frequencies but not based on Y-STR variation within the main subhaplogroups. A clear gradient was found with higher frequencies of R1b1b2 (R-M269) chromosomes in the northern *vs.* southern regions, mainly related to a trend in the frequency of R1b1b2a1 (R-U106).

1. Introduction

Genetic diversity is geographically unequally distributed among human populations. The ancestral origin and evolutionary forces such as selection, drift and migration have played a crucial role in genetic population differentiation [1]. It is required that the geographical distribution of genetic variation is known when genetic tools are being used in forensic science [2]. This is especially the case for the application of the haploid Y-chromosome (Y-chr) due to the high effects of genetic drift and to the strong susceptibility of founder events for this chromosome [3,4]. In addition, the occurrence of patrilocality in approximately 70% of the modern human societies increases the micro-geographical clustering of the Y-chr variation in comparison with mitochondrial variation [5]. Currently, the Y-chr distribution is well known on a continental scale, nevertheless, one of the future issues will be to study the regional distribution of the genetic variation [1,6].

Research on a regional or micro-geographical scale requires attention to essential issues such as an intensive sampling campaign and a fine resolution detection of Y-chr variation to differentiate unrelated families [1]. During sampling, most population studies of Y-chr diversity classify donors into local subpopulations on the basis of at least two generations of residence [7]. However, this is compromised by migration in preceding generations, especially in Western Europe since the beginning of the 19th century with the industrial revolution. It is therefore essential to know the genealogical context of each donor for many generations to study regional population structure.

The framework to study population stratification on a micro-geographical scale for Y-chr was optimised for a selected central-western European region named Brabant. The Duchy of Brabant was a historical region in the Low Countries between the 12th and 18th century and consisted of a present-day Dutch province and three contemporary Belgian provinces together with the Brussels-

* Corresponding author at: Laboratory of Forensic Genetics and Molecular Archaeology, Kapucijnenvoer 33, B-3000 Leuven, Belgium. Tel.: +32 16 33 66 63; fax: +32 16 34 59 97.
E-mail address: maarten.larmuseau@bio.kuleuven.be (M.H.D. Larmuseau).

1872-4973/$ – see front matter © 2010 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2010.08.020

Capital Region. The total area is 14.425 km^2 with approximately 150 km between the two most remote places in Brabant. The main reason for selecting this region was the ability to obtain reliable genealogical data of the patrilineal line for each of the numerous donors living together on a small geographical scale. This provided an optimal starting point to study micro-geographic distribution of Y-chr variation in Western Europe.

2. Materials and methods

Buccal swab samples were collected from a total of 477 males representing 423 different surnames. Only males that provided genealogical data of the patrilineal line with at least one known ancestor living in the 18th century were selected for this study. According to the residence of the oldest known parental ancestor, each donor was assigned to one of the five 'genealogical regions' within Brabant based on contemporary administrative borders (Noord-Brabant, Antwerpen, Kempen, Mechelen and Vlaams- and Waals-Brabant; Fig. S1). DNA was extracted by using the Maxwell® 16 System (Promega, Madison, USA) and quantified by real-time PCR (Quantifiler™ Human DNA kit, Applied Biosystems).

In total 37 STR loci were genotyped for all samples as described in a previous study [8] based on PowerPlex® Y (Promega, Madison, USA) (DYS391, DYS389-I, DYS439, DYS389-II, DYS438, DYS437, DYS19, DYS392, DYS393, DYS390, DYS385) and three novel multiplexes (DYS426, DYS393, DYS390, DYS385, DYS460, GATA H4.1, DYS447, DYS448, DYS459, DYS576, DYS464, YCAII, DYS456, DYS458, DYS607, DYS455, DYS570, DYS724, DYS454, DYS388, DYS442). The inclusion of DYS464 into two assays facilitated the interpretation of the alleles and peak height ratios [9]. In addition, some STRs were included in more than one multiplex to serve as an internal control. The whole process was reproduced with new primer sets for all individuals that showed non-amplified loci to exclude technical errors or mutations in the standard primer positions.

All haplotypes were submitted to Whit Atheys' Haplogroup Predictor (Athey 2005; Athey 2006) to obtain probabilities for the inferred haplogroups. This strategy was used to avoid redundant SNP-typing, though, verification of the haplogroup with Y-SNPs was required [10]. Based on these results, the samples were assigned to a specific SNP assay to confirm the haplogroup and to assign the subhaplogroup to the lowest possible level of the latest Y-chr tree reported by Karafet et al. [11] and according to the update on the Y Chromosome Consortium web page (http://ycc.biosci.arizona.edu/nomenclature_system/index.html), with exception of the substructuring within subhaplogroups R1b1b2a1 (R-U106) and R1b1b2a2g (R-U152). Fifteen multiplex systems with Y-SNPs were developed using SNaPshot mini-sequencing assays (Applied Biosystems, Foster City, CA) and analyzed on an ABI3130XL Genetic Analyzer (Applied Biosystems) according to a previously published protocol [12]. Some Y-SNPs were analysed by sequencing using the BigDye Terminator v. 3.1 (Applied Biosystems) or by allele-specific-amplification using SYBR green with the 7500 real-time PCR system (Applied Biosystems). All primer sequences and concentrations for the analysis of the 103 Y-SNPs are available from the authors upon request.

The genetic relationship between different populations was assessed by means of Φ_{ST}, an analogue of Wright's F_{ST} that takes the evolutionary distance between individual haplotypes into account [13]. Estimations of Φ_{ST} were calculated based on the Y-SNP subhaplogroup frequencies and on the 25 single-copy Y-STRs (including 'DYS389-1' instead of DYS389-I and DYS389-2, which is DYS389-II–DYS389-I) between all regions, as well as between a single region and all the other regions combined. To calculate the genetic relationship between populations based on microsatellite data also the R_{ST}, another analogue of the F_{ST}, was used which takes the difference in repeat numbers between alleles into account [13]. R_{ST}-values, estimated as ρ [14], were calculated based on the Y-STR data between all regions as well as between a single region and all the other regions combined. Φ_{ST} and R_{ST} estimates were also calculated based on Y-STR data within the two most frequent observed subhaplogroups R1b1b2a1 (R-U106) and R1b1b2a2* (R-R312*). All Φ_{ST}- and R_{ST}-values were obtained by taking only one participant into account for pairs with the same family name, the same 'genealogical region' and belonging to the same subhaplogroup, to exclude the possibility of family effect in the analysis. All values were estimated using ARLEQUIN v.3.1 [15] and tested for statistical significance by means of random permutation of samples in 10,000 replicates. For the pairwise Φ_{ST}- and R_{ST}-values, the sequential Bonferroni correction was applied to correct significance levels for multiple testing [16].

Median joining networks for all haplogroups and the main subhaplogroups were constructed based on all 25 single-copy Y-STRs by NETWORK 4.5.1.0. [17] (http://www.fluxus-engineering.com) using the weighting scheme described by Qamar et al. [18] due to different mutation rates among the markers. To estimate the time to the most recent common ancestor (tMRCA) of the main subhaplogroups, we used all 25 single-copy Y-STRs and applied the average square distance (ASD) method [19], where the ancestral haplotype was assumed to be the haplotype carrying the most frequent allele at each microsatellite locus. We employed a microsatellite evolutionary effective mutation rate based on the observed father-to-son transmissions of all used microsatellites according to Vermeulen et al. [2] and using the correction of Zhivotovsky et al. [20]. The tMRCA estimates and confidence intervals (CI) were calculated with the software Ytime v.2.08 [21].

3. Results

3.1. Y-chromosomal variation

All individuals were correctly assigned to the main haplogroups using the Whit Atheys' Haplogroup Predictor. In total, eight main haplogroups were observed with almost 85% of the samples belonging to haplogroup R (63%) and I (21%) (Table 1). On the lowest observed level of the phylogenetic tree 32 subhaplogroups were found in the data set, whereby nearly 70% of all samples belonged to only four subhaplogroups: R1b1b2a1 (R-U106), R1b1b2a2* (R-P312*), R1b1b2a2g (R-U152) and I1* (I-M253*) (Table 1).

For the 477 males, a total of 286 different 'minimal haplotypes' (=DYS19, DYS389-1, DYS389-2, DYS390, DYS391, DYS392, DYS393 and DYS385a,b) were observed, of which 209 were unique. The most frequent 'minimal haplotype' occurred 33 times (7%); the frequencies of all 77 'minimal haplotypes' that were observed more than once in the dataset are given in Table S1. A total of 337 different 'extended haplotypes' (='minimal haplotype' + DYS438 and DYS439) were observed, of which 271 were unique. The two most frequent 'extended haplotypes' occurred both 18 times (3.8%); the frequencies of all 66 'extended haplotypes' that were observed more than once in the dataset are given in Table S2. Many similar 'minimal and extended haplotypes' belonged to individuals that were assigned to a different subhaplogroep based on Y-SNPs. Using all 37 Y-STRs, 473 haplotypes were observed in the study, of which 469 were unique. The four duos with the same 37-STR haplotype also had an identical surname. All 'extended haplotypes' together with the SNP-typing results have been submitted to the Y-STR Haplotype Reference Database (www.yhrd.org; Accession numbers YA003651-YA003652-YA003653).

Network analyses of all single-allele Y-STR haplotypes within the main haplogroups was able to differentiate the Y-SNP defined subhaplogroups from each other, except for the subhaplogroups