# ORIGINAL ARTICLE High Y-chromosomal diversity and low relatedness between paternal lineages on a communal scale in the Western European Low Countries during the surname establishment

MHD Larmuseau<sup>1,2,3</sup>, N Boon<sup>4,5</sup>, N Vanderheyden<sup>1</sup>, A Van Geystelen<sup>3</sup>, HFM Larmuseau<sup>6</sup>, K Matthys<sup>6</sup>, W De Clercq<sup>7</sup> and R Decorte<sup>1,2</sup>

There is limited knowledge on the biological relatedness between citizens and on the demographical dynamics within villages, towns and cities in pre-17th century Western Europe. By combining Y-chromosomal genotypes, in-depth genealogies and surname data in a strict genetic genealogical approach, it is possible to provide insights into the genetic diversity and the relatedness between indigenous paternal lineages within a particular community at the time of the surname adoption. To obtain these insights, six Flemish communities were selected in this study based on the differences in geography and historical development. After rigorous selection of appropriate DNA donors, low relatedness between Y chromosomes of different surnames was found within each community, although there is co-occurrence of these surnames in each community since the start of the surname adoption between the 14th and 15th century. Next, the high communal diversity in Y-chromosomal lineages was comparable with the regional diversity across Flanders at that time. Moreover, clinal distributions of particular Y-chromosomal lineages between the communities were observed according to the clinal distributions earlier observed across the Flemish regions and Western Europe. No significant indication for genetic differences between communities with distinct historical development was found in the analysis. These genetic results provide relevant information for studies in historical sciences, archaeology, forensic genetics and genealogy.

Heredity (2015) 115, 3-12; doi:10.1038/hdy.2015.5; published online 15 April 2015

### INTRODUCTION

One's community has always had an important role in defining aspects of an individual's identity. Therefore, historians and social scientists have been studying for decades the demographic history and kinship between citizens within settlements in Western Europe from the origin of the village, town or city-further referred as 'community'-until present day (Schürer, 2004). In this context, archaeology including physical anthropology may provide data on the start of the settlement, as well as on the geographical and demographical evolution of a community. Archival documents provide added value to the historical background of a community, as well as the estimates of community sizes based on the past censuses and vital statistics (Willigan and Lynch, 1982). Moreover, archives are the main source of information on familial relatedness and individual dispersion events as they allow the study of surnames and genealogical sources in particular. Surnames are interesting because they have been patrilineally inherited since the 13th century and commonly used in the 1500s in several Western European regions (King and Jobling, 2009b). Next, in-depth and population-wide genealogical research can be performed in Western Europe from the end of the 16th century onwards because parish registers and civil records can be consulted (Willigan and Lynch, 1982). It is difficult, however, to link archaeological and historical data of a particular community with the first occurrences of families with surnames in that community. Therefore, it is hard to get insight into the demographic evolution within a community and into the biological relatedness between citizens of a community before modern history (<1600). Nevertheless, genetic analyses are promising to provide data for filling this research gap in the historical survey of West-European communities.

Genetic data can provide important insights into the demographical continuity and kinship within a community by using both ancient DNA and modern DNA approaches. Firstly, ancient DNA analysis on archaeological material may reveal the genetic diversity at a certain location and at a particular time. It is, however, difficult to obtain enough samples to make statistically relevant conclusions for a population, especially owing to the practical difficulties of retrieving verifiable and contamination-free DNA data and the lack of sufficient individuals (Larmuseau *et al.*, 2013b). Secondly, DNA of currently living individuals may suggest past relatedness and genetic diversity within the population of a village or region (Winney *et al.*, 2012).

<sup>&</sup>lt;sup>1</sup>Laboratory of Forensic Genetics and Molecular Archaeology, UZ Leuven, Leuven, Belgium; <sup>2</sup>Department of Imaging and Pathology, Forensic Biomedical Sciences, KU Leuven, Leuven, Belgium; <sup>3</sup>Laboratory of Socioecology and Social Evolution, Department of Biology, KU Leuven, Leuven, Belgium; <sup>4</sup>Laboratory of Biodiversity and Evolutionary Genomics, Department of Biology, KU Leuven, Leuven, Belgium; <sup>5</sup>Institute of Tropical Medicine, Antwerp, Belgium; <sup>6</sup>Faculty of Social Sciences, Department of Social Sciences, Centre of Sociological Research (CESO), KU Leuven, Leuven, Belgium and <sup>7</sup>Faculty of Arts and Philosophy, Department of Archaeology, Ghent University, Ghent, Belgium Correspondence: Dr MHD Larmuseau, Department of Imaging and Pathology, Forensic Biomedical Sciences, KU Leuven, Kapucijnenvoer 33, Leuven B–3000, Belgium.

Received 10 July 2014; revised 20 November 2014; accepted 23 December 2014; published online 15 April 2015



Figure 1 Map of the study area. The coloured area in light grey represents Flanders (including the selected adjacent parts of the Dutch provinces—Zeeland and Limburg); the areas coloured in darker grey are the regions northern West-Flanders (NW-Flanders), southern East-Flanders (SE-Flanders), southern Brabant (S-Brabant) and southern Limburg (S-Limburg); the black points are the selected communities with Oudenburg (O), Snellegem (S), Velzeke (V), Idegem (I), Alken (A) and Tongeren (T).

However, modern DNA sampling in Western Europe-even when the birthplaces of grandparent's are taken into account to collect DNA donors-will always provide a blurred and misleading picture of a specific past time period under study owing to more recent migrations and expansions (Winney et al., 2012; Larmuseau et al., 2013b). To deal with this problem, the unique link between a heritable cultural marker -the patrilineal surname-and a genetic marker-the Y chromosome -provides an opportunity to target sets of living individuals that might resemble populations at the time of surname establishment until today (Bowden et al., 2008; Larmuseau et al., 2012b). In addition, it is possible to organise sampling campaigns for a specific location by using the genetic genealogical approach, which selects indigenous surnames based on the historical documents and carries out in-depth genealogical research for each DNA donor to exclude descendants of illegitimate children, adoptees, extra-pair paternity (EPP) and migrants with an adopted local surname (Larmuseau et al., 2013b).

The genetic genealogical approach at a communal level has a high potential value for historical surveys on the biological relatedness between paternal lineages and the genetic diversity at the time of surname establishment. Although, a study on (West-European) communities with the required criteria of the genetic genealogical approach has not been performed yet to our knowledge. This is mainly because of the time-consuming sampling criteria (Larmuseau *et al.*, 2013b). An optimal Western European region to study the value of this approach on a community level is Flanders (Belgium) owing to

Heredity

its central location, its lack of noteworthy cultural and geographical isolates in the 'open' geography, and the decades-long tradition of performing archaeology, historical population registration and surname studies on a communal and regional scale (Cloet and Vandenbroeke, 1989; Debrabandere, 2003; Barrai et al., 2004). Moreover, the Y-chromosomal diversity at a high phylogenetic resolution is already well known on a regional scale in Flanders which revealed a well-studied historical genetic pattern (Larmuseau et al., 2012a, 2014b). Also a stable low EPP rate of 1-2% per generation in the last 400 years is observed within Flanders, making the genetic genealogical approach feasible in this region (Larmuseau et al., 2013a). Next to optimizing and evaluating the approach on a communal scale, two concrete research questions will be answered by using this methodology within six particular Flemish communities which were selected based on their differences in geography and local history: (i) 'Was the biological relatedness between indigenous patrilineages greater within communities than within regions at the time of the surname establishment?' and (ii) 'How different is the Y-chromosomal diversity in indigenous patrilineages from communities versus regions? And are the genetic differences between communities mainly due to genetic drift or due to variation in past gene-flow events or historical development?'. Therefore, this study is not aimed at directly relating identity and ethnicity as these are complex and multi-layered social constructions (Jones, 1997). Rather kinship in paternal lineage and its relation to geography is the subject of debate and enquiry in this study.

Finally, the results will be discussed in the light of its applications in historical sciences, socio-demography, archaeology, anthroponomy, genealogy and forensic genetics.

## MATERIALS AND METHODS

## Ethics statement

This study has been approved by the institutional review board, namely the UZ Leuven Medical Ethics Committee, under protocol number S54010.

## Selection of communities and regions

Six villages and towns, further referred as 'communities', were selected within contemporary Flanders based on their geography and historical development (Figure 1). The six communities are classified geographically in three pairs with a pair by the coast in the province West-Flanders, namely Oudenburg and Snellegem; a pair in central Flanders in the province East-Flanders, namely Velzeke and Idegem; and a pair in the most eastern part of Flanders in the province Limburg, namely Tongeren and Alken. Within each of these three pairs, one locality is known to be populated since the Roman period (58 BC-circa 410 AD; further referred to as the 'Gallo-Roman' or 'GR' research group), the other locality is known to be a settlement which mainly developed since the Early Middle Ages (further referred to as the 'Early medieval' or 'EM' research group; see Supplementary Materials). This does of course not mean that the communities within the EM research group were not populated before the Early Middle Ages, it means that the historical development of the communities mainly started after the Roman period. The influence of historical development in genetic diversity between communities may result in differences between the GR and EM groups, as several previous large-scale genetic studies have assigned distributions of some specific nuclear and Y-chromosomal variants in Western Europe to migrations already during the Roman Empire, although younger Germanic migration events gradually reshaped this region during the decline of the Roman Empire (King et al., 2007; Faure and Royer-Carenzi, 2008).

Four regions were defined to compare the Y-chromosomal diversity and biological relatedness within indigenous patrilineages between the communal versus regional scale at the time of the establishment of surnames. Three regions were defined by a circular area with a radius of 30 km and with the exact position between the two communities within each geographical pair as the middle point (Figure 1). The radius of 30 km for a region was selected as this is the typical distance a person can walk in one day on a (relatively) plain landscape and it is also the averaged distance between main towns in the Low Countries. On the basis of this rule, we defined a NW-Flanders region, a SE-Flanders region and a S-Limburg region (Figure 1). In addition, we defined an extra region between the East-Flanders and the Limburg regions, namely the S-Brabant region, with also a radius of 30 km and with the town Leuven as middle point. With this extra region it was possible to study the full West-East axis within Flanders (Figure 1). Finally, the overall region 'Flanders' was defined as the sum of all four regions and the rest of contemporary Flanders including the parts of the Dutch provinces Zeeland and Limburg which are adjacent to, and have a strong historical connection with Flanders (Figure 1).

## Sampling procedure

The selection of the DNA donors was crucial to provide good reference samples for the populations in the six communities to answer both research questions. This rigorous selection was based on the strict criteria of the genetic genealogical approach providing an authentic sample for a certain area at the time of the surname establishment (Larmuseau *et al.*, 2013b). Firstly, a list of selected surnames for each community was collected based on the extended archival research. All surnames which are known to have occurred from the first announcements of surnames starting from the 14th century until the year 1575 in the community were listed based on the archival material (see Supplementary Materials for sources). All surnames with an origin, toponym or an indication of language or dialect from an area outside the selected community based on the anthroponomical research (Debrabandere, 2003) were erased from these lists. No differences in the amount of archival resources were found between communities and regions. A higher number of selected surnames were collected for Oudenburg and Tongeren (25% more names), as both cities had a higher census population size during the Middle Ages in comparison with the other four selected communities. Secondly, huge efforts assisted by national and regional media, archives, municipalities and many local volunteers were performed to find potential DNA donors. To be a DNA donor, it was a prerequisite to have a selected surname for a community and their oldest reported paternal ancestor (ORPA) had to be born in the area within a distance of maximally 5 km from the centre of that particular village or town before the year 1800 and preferably before 1700. Descendants of known foundlings, adoptees and unmarried women who passed their surname on were not selected. If possible, two DNA donors of one family lineage were selected to test the occurrence of EPP events in the genealogy of the donors. DNA donors of one family lineage with different Y-chromosome profiles were both excluded from further analyses. This is in contrast to DNA donors with a similar surname (or surname variant) and with different Y-chromosome profiles, if there are no indications for a common known paternal ancestor between both DNA donors according to the extensive archival research. Donors of the same familial lineage were at least related in seventh degree to avoid ethical issues and familial conflicts. In total, 296 DNA donors of the six communities were selected for the genetic analysis.

Samples for the four defined regions and the rest of Flanders were selected from the genetic genealogical databank with currently 1118 samples (Larmuseau *et al.*, 2014b). The same rigorous criteria as for the community samples were used to include DNA donors into one of the region samples or into the overall Flanders sample, keeping in mind that all surnames of the DNA donors have to be present in archival data before 1575 in one of the regions or in Flanders (see Supplementary Table S1 for all selecting criteria of the sampling campaign).

#### Y-chromosome genotyping

The 42 Y-STR haplotype and Y-chromosomal sub-haplogroup were genotyped for each selected participant conforming the method described in Supplementary Material. The 42 genotyped Y-STRs include the 23 loci of the PowerPlex Y23 System (Promega, Madison, WI, USA; Purps *et al.*, 2014) and 19 loci of in-house developed multiplexes (Larmuseau *et al.*, 2011). The Y-chromosomal data genotyped for all 296 donors of the communal samples have been submitted to the open access Y-STR Haplotype Reference Database (YHRD, www.yhrd.org): accession numbers YA003739, YA003740 and YA003742. For the regional populations, most samples were already genotyped for the Y chromosome in previous studies (Larmuseau *et al.*, 2012a, 2014b); this data is accessible on the YHRD (www.yhrd.org) with accession numbers YA003651, YA003652, YA003653, YA003738, YA003739, YA003740, YA003741 and YA003742.

#### Statistical analysis

First, the reconstructed in-depth patrilineages of all selected DNA donors were entered in the genealogical program ALDFAER v. 4.2 (Stichting Aldfaer, 2013; www.aldfaer.net). As such, DNA donors with a known common paternal ancestor were detected. For each of these couples the Y chromosomes were compared between the individuals to verify if their genealogical common ancestor (GCA) was also their biological common ancestor (BCA) according to Larmuseau *et al.* (2013a).

Second, GenAlEx version 6.5 (Peakall and Smouse, 2012) was used to find the sample pairs matching for the set of 38 genotyped Y-STR loci, as this is the set of loci genotyped for the communal, as well as for the regional samples in the analysis although 42 Y-STRs were genotyped for the communal samples. Sub-haplogroup affiliation and the in-depth genealogy of the donors of sample pairs with similar haplotypes were compared with each other. We defined two DNA donors as patrilineal relatives in a historical timeframe when they belong to the same sub-haplogroup and when they have non-matching alleles on maximally 7 out of 38 Y-STRs. This maximum of non-matching loci is proposed based on the calculated mean mutation rate of the 38 Y-STR set using the individual mutation rates measured in Ballantyne *et al.* (2010), namely  $5.91 \times 10^{-3}$  mutations per generation. On the basis of the formulae of Walsh (2001), seven mutations on the 38 Y-STRs would mean that the biological ancestor of both individuals lived between 7 and 36 generations ago (95% credibility interval), that is, between the years 1110 and 1835 if we use 6

a generation span of 25 years or between the years 750 and 1765 if we use a generation span of 35 years. Moreover, the chosen limit of differentiated Y-STRs to declare relatives in a historical timeframe is also supported by the substantial occurrence of a high resemblance in Y-STR haplotypes between males belonging to different sub-haplogroups of the most frequent haplogroup R-M269 and therefore also between males belonging to the same subhaplogroup within R-M269 but without a GCA in historical time (Larmuseau et al., 2014a). In the list of all positive matches between the DNA donors (less than or equal to seven different Y-STR loci), all pairs with the same surname or a spelling variant were reported and only one individual for each pair was used for further analyses. Next, differences in the rate of positive matches (that is, number of positive matches to the total number of combinations) between each DNA donor of a community and another donor of the same community, between each DNA donor of a community and another donor of the region to which the community belongs (including the DNA donors of the other community of the region but excluding the donors of its own community) and between each DNA donor of Flanders (excluding between DNA donors of the same region) were calculated (see Supplementary Materials for a graphical illustration of this analysis). Next to defining a maximum number of Y-STR mutations between two Y chromosomes declaring kinship between two donors on a genealogical time scale, an alternative method was required to verify if the relatedness was different between selected DNA donors of one community, of one region and of Flanders. Therefore, we measured the mean and variance of the variable number of mutated Y-STR loci between each pair of DNA donors belonging to the same Y-chromosomal subhaplogroup in a community. Next, we compared these results with the mean and variance of the variable number of mutated loci between each pair of DNA donors belonging to the same sub-haplogroup in a region and in Flanders. This was only measured and compared for the sub-haplogroups with a frequency of >5% in Flanders.

Third, rarefaction methods allow for a meaningful standardization and comparison of data sets during the quantification and comparison of Y-chromosomal diversity (Gotelli and Colwell, 2001). Rarefaction curves represent the means of repeated re-sampling of all pooled individuals of a particular population (Gotelli and Colwell, 2001). Rarefaction curves for the communities and regions were made based on the 'rarefy' function of the R-package VEGAN (Oksanen *et al.*, 2007). The function 'rarefy' is based on Hurlbert's (1971) formulation of rarefaction, and the calculated standard errors are based on the method explained in Heck *et al.* (1975). In Work *et al.* (2010), a R-script is provided that makes a loop of this VEGAN function resulting in individual-based rarefaction curves including standard errors.

Fourth, sub-haplogroup frequencies were estimated and compared between each defined community and region. Pairwise FST values between the communities and regions were estimated using ARLEQUIN v.3.1 (Excoffier et al., 2005). Significance of population subdivision was tested using a permutation test implemented in R (The R Community, 2011), as developed by Larmuseau et al. (2012b). In the case of the pairwise tests the Bonferroni correction was applied to all the P-values (Rice, 1989). Statistically significant differences of frequencies for frequent sub-haplogroups (global frequency of >5%) between the communities and regions were tested using  $\chi^2$  tests in R. Next, a principal component analysis was performed with R as a clustering analysis of the communities, including a biplot that plots on the same plane the vectors representing the contribution of each of the original variables to these components. A correspondence analysis based on the Y-chromosome haplogroup frequencies (without unique sub-haplogroups) was performed using the correspondence analysis package of R. Next, correlations between pairwise  $F_{ST}$ -values and the geographic distance between the communities were calculated and tested using simple Mantel procedures (Mantel, 1967) in the VEGAN package in R (Oksanen et al., 2007). Because the number of Mantel test permutations is limited for small sample sizes (n = 6), complete enumeration of all possible 6! = 720 permutations of the (first) dissimilarity matrix was carried out for all tests. Finally, as the multivariate analogue of regression, a redundancy analysis (RDA) was performed with R to study the influence of geography versus historical development (the GR and EM research groups) on the distribution of the Y-chromosomal lineages in the six communities.

### RESULTS

The huge interest and participation of the wide public for this study resulted in 296 selected DNA donors from the six communities in accordance to all the criteria of the genetic genealogical approach. The in-depth genealogy of each selected DNA donor was reconstructed; 11% of the donors had an ORPA living before 17th century, 54% in the 17th century and 35% in the 18th century. In this data set nine independent couples of DNA donors with a GCA were observed based on their in-depth genealogies. For all of these couples the GCA was also the BCA, as the individuals within each couple were assigned to the same subhaplogroup at the highest phylogenetic resolution and as their haplotypes revealed no more than 7 Y-STR differences out of 38 Y-STR loci (for results see Supplementary Materials, Supplementary Table S2). To avoid a bias by including several members of the same family, only one individual per genealogical pair with a confirmed BCA was selected for further analysis. Next, the selected 38 Y-STR haplotypes of the communities were compared with the ones from the full updated data set of Larmuseau et al., 2014b, wherefore already the couples with a known GCA were checked for a BCA (Larmuseau et al., 2013c). After the analysis, in total 110 haplotype matches with maximally seven different Y-STR loci were observed between DNA donors with the same (or similar) surname. One donor for each of such match was always excluded from further analysis to avoid a family bias. A match was also found for some DNA donors with a similar surname or a spelling variant but from different communities or regions. There was a Y-chromosomal match between males with three different spelling variants of the same surname which were present in archival records of three different communities (Oudenburg, Idegem and Velzeke) before the 15th century. After these analyses, 253 selected DNA donors for the six communities met all the criteria, next to 418 additional donors for the four described regions (NW-Flanders, SE-Flanders, S-Brabant and S-Limburg) and 256 extra donors for the rest of Flanders (see Supplementary Table S1 for all sampling criteria). An overview of the sub-haplogroup frequencies for the communities and regions is given in Supplementary Table S3 (Supplementary Materials).

After removing the 110 Y-chromosomal matches (≤7 mutated Y-STR loci and same sub-haplogroup) between DNA donors with the same (or similar) surnames, only 19 matches were found between DNA donors assigned to the same community (out of 5376 combinations). Next, 59 (out of 27 166 combinations) and 256 (out of 153 227 combinations) matches were found between a DNA donor assigned to one selected community with a donor assigned to the same region (exclusive to the same community but including the other community of the region) or Flanders (exclusive to the same region), respectively (see Supplementary Materials for a graphical illustration of this analysis). Therefore, the chance for a match between two samples assigned to the same community is  $0.35 \pm 0.16\%$ ; between two samples assigned to the same region (exclusive to the same community) it is  $0.22 \pm 0.05\%$ ; and between two samples assigned to Flanders (exclusive to the same region) it is  $0.14 \pm 0.02\%$ . Finally, no significant differences were observed in the means and variances of the variable number of mutated Y-STR loci between each pair of DNA donors in a community, region and Flanders which belong to the same sub-haplogroup (*P*-values > 0.05; results not shown). These were measured for all DNA donors which were assigned to subhaplogroups I-M253\*, R-L2\*, R-L48, R-P312\* and R-Z381\*, as they had a frequency of >5% in the global Flemish sample (Supplementary Table S3).

The rarefaction curve with the calculated standard errors based on the method of Heck et al. (1975) showed no significant differences



Figure 2 Rarefaction curves of the Y-chromosomal diversity of (a) three regions (NW-Flanders, SE-Flanders and S-Limburg) and all six selected communities, and (b) all four defined regions (NW-Flanders, SE-Flanders, S-Brabant, S-Limburg) inclusive of the six communities.

between the Y-chromosomal diversity within a community versus the diversity within its corresponding region (Figure 2a, Supplementary Figure S3). Two clusters of regions were found based on the Y-chromosomal diversity with the rarefaction curve, namely NW-Flanders and SE-Flanders did not differ from each other, and neither did S-Brabant and S-Limburg. These two clusters differed significantly from each other with a higher diversity for S-Brabant and S-Limburg in comparison with NW-Flanders and SE-Flanders. Clustering of populations based on the Y-chromosomal diversity does not mean that populations included in these clusters are also more similar based on the frequencies of Y-chromosomal haplogroups. The clustering is clearer when the samples of the communities are pooled with the regional samples (Figure 2b). On communal level the same gradient of higher Y-chromosomal diversity from East-to-West is visible (Figures 2a and b). The significant highest diversity is found for Tongeren and Alken, which is comparable with the diversity in their corresponding region S-Limburg. The significant lowest diversity is found for Oudenburg and Idegem, although the diversity in Snellegem and Velzeke is not significantly different from the diversity in their regions NW-Flanders and SE-Flanders.

The sub-haplogroup frequencies for each defined community and region is given in Supplementary Table S3. No significant pairwise

 $F_{ST}$  values between communities and regions were found (P-values > 0.05), except the pairwise  $F_{ST}$  values between Alken versus the four communities of NW-Flanders and SE-Flanders were non-significant after Bonferroni correction (P-values between 0.02 and 0.04). No significant differences of frequencies for the sub-haplogroups between the communities and regions were found (*P*-values of the  $\chi^2$ -tests>0.05). A clear clustering between the communities was found in the principal component analysis plot according to their geography, and especially based on PC1 which explains 61% of the distribution of Y-chromosomal variation (Figure 3). The biplot shows that this geography was mainly based on the frequencies of sub-haplogroups R-L48, R-M529 and I-M253 which were identical to the East-West gradients of the frequencies for these sub-haplogroups on a regional scale (Supplementary Table S3). A similar clustering between the communities according to their geography was as well found in the correspondence analysis plot, and especially based on the first dimension which explains 34.68% of the distribution of Y-chromosomal variation (Figure 4). Next, the Mantel test showed a significant correlation between the FST-values and geographical distances between the six communities (R = 0.5284; P-value: 0.0481). Finally, the RDA was not powerful because of the low number of degrees of freedom; however, indications of a

Heredity



**Figure 3** Principal component analysis (PCA) plot of the six communal samples based on the Y-chromosomal diversity together with a *biplot*; cumulative proportion is 0.74 for the first two principal components (PC1: 0.61; PC2: 0.13).

geographical pattern of the distribution of the Y-chromosomal lineages were found as the constrained variable was marginally nonsignificant in the RDA, instead of influences of the historical development (the GR and EM research groups) of the communities (see Supplementary Materials).

## DISCUSSION

On the basis of the modern DNA samples, in-depth genealogies, extensive surname data and high interest of the wide public for participation in history-oriented genetic studies (Scully *et al.*, 2013), it was possible to link Y-chromosomal variants to citizens from six communities and four regions in Western Europe at the time of the surname establishment between the 14th and 15th century. This survey revealed a similar (i) low relatedness between indigenous paternal lineages and (ii) high Y-chromosomal diversity and low differentiation on the communal versus regional scale.

Low relatedness between indigenous surnames within a community A high divergence was observed in this study between the Y-chromosomal haplotypes of the paternal lineages which have co-occurred on a communal scale since surname establishment. Biological relatedness based on Y-chromosomal comparisons within one single community was almost exclusively observed between DNA donor pairs with a common genealogical ancestor (CGA) and/or with the same or similar surname. Without those pairs with highly expected relatedness, the chance for a haplotype match between two donors assigned to the same community (with differences on maximally 7 Y-STRs out of 38 Y-STR loci) was surprisingly low  $(0.35 \pm 0.16\%)$  and not significantly higher than between donors assigned to the same region but excluding the same community  $(0.22 \pm 0.05\%)$ . On the other hand, the communal and regional frequencies of biological relatedness were marginally but significantly higher than the rate between samples assigned to whole of Flanders (excluding the same region;  $0.14 \pm 0.02\%$ ), showing, as expected, lower chance of finding a



Figure 4 Correspondence analysis (CA) of the six communal samples based on the Y-chromosomal diversity (without unique sub-haplogroups).

patrilineal relative the further you diverge the place of birth of the ORPA (Ralph and Coop, 2013). Although low frequencies of Y-chromosomal haplotype matches on communal and regional scale were observed, the chance for a false positive match between haplotypes owing to identical-by-state is still high because of the recently discovered strong resemblance between haplotypes within the frequent haplogroup R-M269 (Solé-Morata et al., 2014; Larmuseau et al., 2014a). Therefore, the means and variances of the variable number of mutated Y-STR loci between all pairwise combinations of DNA donors within each main sub-haplogroup were also compared between communities and regions but as before, no significant difference was found. Because of the low observed number of Y-chromosomal matches, no dominant Y-chromosomal haplotypes were thus found within communities and regions at the time of surname adoption in the Late Middle Ages (results not shown). This is in contrast to Ireland (Moore et al., 2006) and Central Asia (Zerjal et al., 2003) where regional dominant haplotypes were observed in the present population but which were associated with specific historical facts before modern history (<1600).

The absence of dominant haplotypes and the low relatedness between the families with indigenous surnames in the selected communities indicate in the first place a low past EPP rate in these populations. Several behavioural studies suggested historical EPP rates of 10-30% per generation (Anderson, 2006), but if past EPP rates were indeed that high, matches should be expected to occur frequently in this study as the surnames co-occurred for centuries in the small populations. Although EPP by men that are paternally related to the supposed father will not be detected, the low historical EPP rate is further illustrated by the fact that all nine couples in the communities with a known CGA were indeed patrilineally related based on their Y chromosomes. In addition, as expected in populations with low past EPP rates, the relatedness between DNA donors with the same surname or a spelling variant over the whole data set of this study was dependent on the frequency of the surname in Flanders (results not shown). This has been extensively shown for the population in England where different Y-chromosomal lineages are still observable in highly frequent surnames owing to independent origins of a frequent surname, suggesting that the past population was characterized with a low EPP rate (King and Jobling, 2009a). A recent study reported indeed already a past EPP rate of 1-2% in Flanders over the last few centuries (Larmuseau *et al.*, 2013c). Owing to the low historical EPP rate and the control of genealogical records for each DNA donor to avoid non-patrilineally inherited surname adoptions, we may assume that the low number of Y-chromosomal matches observed within the communities reflects the situation at time of surname establishment (Larmuseau *et al.*, 2013b).

Although a high level of genetic drift-in genealogical terms called 'daughtering out' (Helgason et al., 2003; King and Jobling, 2009b)was observed by the extinction of many surnames that had persisted for many generations within each community, the low kinship between the survived surnames indicates that one or several (subsequent) migration events had occurred changing the population within the communities during the period of surname establishment. The observation of low relatedness between indigenous families is surprising as synchronic analyses for each time point in the last 400 years showed patrilocality and huge family networks in Western European communities whereby almost every citizen had at least one familial connection in the community itself (Cloet and Vandenbroeke, 1989; Fincher and Thornhill, 2008). The adoption of surnames in Western Europe was after all a long-term process lasting several centuries so that several potential events may have caused the inferred high dispersion rate. One important migration event in Western Europe coincides with the peak of surname establishment, namely the bubonic plague or the 'Black Death'. This pandemic disaster occurred circa 1350 and caused a huge depopulation by death and migration, even in rural areas (Redmonds et al., 2011; DeWitte, 2014). Such huge migration event most likely resulted in a shake-up of the already adopted surnames in parishes so that at the moment of the first notions of surnames in archive documents of a certain community, many surnames were no longer close to their points of origin (Redmonds et al., 2011). Next to a shake-up of paternal lineages in a parish, one hypothesis stated earlier that hereditable surnames were precisely established after migration or in periods of high migration rates because of several reasons (Redmonds et al., 2011): as migrants received surnames in their new community which reflect their origin by a toponym; or as an inhabitant wanted to stress their inheritance in a period of many immigrants by a surname which refer to their patrilineal descent with a patronym; or when immigrants and traders introduced the use of surnames as a status symbol or requirement for trading to the inhabitants, a process which is documented in the Netherlands where surnames only started to be common once the Flemish migrants and other traders with surnames migrated to Holland in the 17th century (Debrabandere, 2003). Although migration may have had indeed a substantial influence in the establishment of surnames, the process whereby each family in a West-European population used a patrilineal inherited surname was anyway a complex one with several factors involved (Redmonds et al., 2011).

# High Y-chromosomal diversity and low differentiation on communal scale

Apart from the large differences between Y-chromosomal haplotypes on communal scale, a high Y-chromosomal diversity of evolutionary lineages was observed in the samples of the six selected communities which reconstruct the diversity at the time of the surname adoption in the Late Middle Ages (Supplementary Table S3). Although the already mentioned extinction of many paternal lineages because of the genetic drift, the observed high Y-chromosomal sub-haplogroup diversity within the six communities was comparable with the diversity on a regional scale. The same pattern of sub-haplogroup diversity with a significant higher diversity in the Western versus the Eastern regions in Flanders was even found on the community level, as the Y-chromosomal diversity in the four communities of the two most western regions was significantly lower than that within the two communities of the most eastern region in Flanders (Figure 2a). Therefore, communal Y-chromosomal diversity appears to not be strongly influenced by genetic drift at the time of surname adoption. These results also indicate no measurable influence of differences in demographic histories between the communities based on the Y-chromosomal diversity within the indigenous patrilines, although Oudenburg and Tongeren had a higher population size during the Middle Ages in comparison with the other selected communities. Of course owing to the adopted sampling procedure the effects of genetic drift within the communal population since the time of surname adoption are not studied. The low effect of genetic drift on the Y-chromosomal diversity at communal level at the time of surname adoption is also visible in the persistence of the East-West gradients in terms of the frequencies of two R-M269 lineages and the I-M253 sub-haplogroup (Figure 3) which were previously observed at a regional scale within Flanders based on the indigenous patrilines (Larmuseau et al., 2014b). Remarkably, these East-West gradients were still visible on regional scale within Flanders when the samples were recruited according to the standard grandparent's criterion (Larmuseau et al., 2014b). Moreover, these East-West gradients in terms of the frequencies of the mentioned sub-haplogroups were also observed based on all present patrilines in the Low Countries (P de Knijff and E Altena, unpublished data) and Europe (Cruciani et al., 2011; Busby et al., 2012).

The observed East-West Y-chromosomal diversity gradient in Flanders on communal and regional scales based on the indigenous patrilines (Figures 2a and b and Supplementary Figure S3) is remarkable as the most western regions and communities (Snellegem and Oudenburg) are located along the coast. A higher genetic diversity at the time of the surname adoption in the Late Middle Ages was assumed within local coastal populations owing to a higher expected rate of historical emigration and trade transport from distant areas along the 'open' sea line, which were definitely high during the Roman Empire and the Early and later Middle Ages based on the historical research and several archaeological studies on material culture (Loveluck and Tys, 2006). The higher observed diversity in the eastern part versus western part of Flanders is therefore unexpected, especially with the knowledge that cities as Bruges and Ghent in the western part of Flanders were leading economic and cultural centres in Europe during the Middle Ages with for instance permanent colonies of Mediterranean merchants already present from the 13th century onwards in Bruges (Hillewaert et al., 2011). On the other hand, during the Roman Empire the eastern region of Flanders was closely connected by road with the Rhineland, which was a heavily Romanized and urbanized area by the presence of the Roman army and of several large towns (coloniae such as Köln and Xanten; Bechert, 2007). Moreover, trade and migration by major rivers-such as the Maas-during the early middle ages might as well have augmented the genetic diversity in the eastern region of Flanders. How a population diversity pattern is created is a complex issue and such a pattern has to be viewed as a palimpsest in which multiple demographic events from different periods are superimposed (Jobling, 2012). Nevertheless, there is no indication that the difference in genetic diversity between eastern and western Flanders should be linked with differences in the amount of available archival sources between Flemish regions, as it was already noticed during the archival research and was expected based on the decades-long historical demographical research of this area (Cloet and Vandenbroeke, 1989).

10

It is tempting to link specific genetic diversity and a population genetic pattern to one well known historical period (Jobling, 2012). To avoid such a 'historical cherry-picking', the six particular communities in this study were selected to survey in a statistically rigorous way if a genetic impact of a historical development within a community during the Roman Empire versus the Early Middle Ages in Flanders is still observable using the modern DNA samples on communal scale. On the basis of the material-culture studies this binary view has been challenged intensively, as Roman-Germanic exchanges were intensive already during Roman times, making too binary-opposed interpretations on aspects of identity (such as ethnicity or status) after the Roman times problematic (Halsall, 2007). On the other hand, several previous studies assigned distributions of some specific nuclear and Y-chromosomal variants in Western Europe to migration during the Roman Period (King et al., 2007; Faure and Royer-Carenzi, 2008). Here, it was tested if the Y-chromosomal diversity within the selected communities was distinguishable based on their historical development including those who already settled during the Roman Empire, as well as those who settled mainly after the Germanic invasions in the early Middle Ages. By using the RDA analysis (see Supplementary Materials), however, no significant indication of differentiation between the two types of Flemish communities in our sampling approach was found. Differentiation between communities was only formed based on the geography, indicating that the bipolar classification of communities in the GR and EM research groups within Flanders has no significance. The lack of a genetic signature of the specific historical development may suggests that contacts between the Roman and Germanic areas were already intensive before the political end of the Western Roman Empire and that during the decline of the Roman Empire, German groups continued gradually to move south and assimilated homogeneously into whole of Flanders (Lamarcq and Rogge, 1996). Otherwise, if there was after all a genetic signature of the historical development, this must have been faded away at the time of surname establishment on a communal scale owing to the entropy or more recent population-wide migration events (Larmuseau et al., 2012a). Therefore, no indication for a historical admixture event during the Roman Period was observed in this Flemish study to clarify the observed high Y-chromosomal diversity within the communities. This is in contrast to a previous study of indigenous patrilines within Northwestern English communities using a similar surname-based sampling approach, where evidence was found for a historical geneflow event from Scandinavia to Northwest England during the Middle Ages (Bowden et al., 2008).

## Conclusions, applications and future issues

This first study using the genetic genealogical approach on communal scale demonstrated the value of this methodology to provide a link between historical and genealogical data in the survey of West-European communities. It could prove that although local patrilocality was common in the last centuries in Western Europe, a high diversity in Y-chromosomal lineages was present next to low relatedness between indigenous paternal surnames which co-occur in a community since the first occurrences of surnames. Moreover, as expected when low genetic drift and substantial gene-flow occurred (Jobling et al., 2013), the Y-chromosomal diversity within a community and the genetic differentiation between communities based on Y chromosomes were related to their geographic position across Flanders at the period of surname establishment. In addition, the same geographical clines of frequencies in particular sub-haplogroups on regional and international scale were also visible on communal scale. The observed clinal distribution of the Y-chromosomal diversity

despite archaeological and historical evidence for genetic discontinuities in Western Europe suggests that future human population genetic studies—especially on a micro-geographic scale—always have to pay attention to recent demographic history in interpreting genetic clines in the light of prehistoric events. This is a conclusion which was already made based on the genetic population structure in the Netherlands by genome-wide analyses (Lao *et al.*, 2013).

Apart from these insights within population genetics, history and archaeology, the results of this study are also relevant for several other research disciplines. Firstly, the highly observed diversity of Y-chromosomal haplotypes and sub-haplogroups within one rural community in Western Europe without strong cultural and geographical barriers is highly informative for forensic genetic casework. Although a higher number of uninformative haplotype matches is expected in rural regions in comparison with industrial regions, the discrimination power between Y chromosomes within rural areas is high enough when appropriate Y-SNP and Y-STR sets are being used. Therefore, Y-chromosomal markers will have enough discrimination power to perform huge DNA surveys on a communal scale, to predict surnames and to include in-depth family searching in forensic casework (Kayser and Ballantyne, 2014). Secondly, the results of our communal survey also confirm the relevance of diachronic analysis within a community or region based on the surname distribution data for historical demographic research (Barrai et al., 2004; Schürer, 2004; Redmonds et al., 2011; Boattini and Pettener, 2013), as a surname might represent one specific Y-chromosomal variation and therefore one specific biological identity in the community. Finally, the study is interesting for each genetic genealogist who is focusing on biological relatedness between their paternal lineage with other families without a known CGA or a similar surname (King and Jobling, 2009a; Scully et al., 2013). The results suggest that within Flanders they do not obtain a higher advantage in their research by performing a sampling campaign in the place of birth of their ORPA in comparison with a region wide sampling.

To better understand processes of communal versus regional Y-chromosomal diversity and relatedness, future research will have to analyse the Y-chromosomal diversity in more regions and communities using the genetic genealogical approach (Larmuseau et al., 2013b). Especially in regions where the process of surname establishment was different in time and length compared with Flanders, the approach will provide new insights. Nevertheless, this approach needs to be compared with genetic research to archaeological data as the genetic genealogical approach only provides an indirect temporal sample limited to individuals who had progeny till today and such a sample will therefore not necessarily represent the whole population at a certain point in the past (Larmuseau et al., 2013b). Moreover, future research has to compare the genetic patterns between samples of the same communities where the DNA donors were recruited according to the strict genetic genealogical approach versus to the standard grandparent's criterion. As such it will be possible to study the differences and similarities in relatedness between residents and genetic diversity within a community since the surname adoption. Next, by comparing data of the genetic genealogical approach on communal and regional scale with ancient DNA samples, it will also be possible to study the influence of genetic drift on the Y-chromosomal diversity (Helgason et al., 2003), as well as natural selection which may occur on the Y chromosome (Wilson Sayres et al., 2014). Preliminary data of an ancient DNA study in the close neighbourhood of our sampling locations, namely in Eindhoven (North Brabant, The Netherlands), based on samples which are dated between 14th and 18th century, revealed a similar high

Y-chromosomal diversity in evolutionary lineages at each defined time period (Altena et al., 2013). Although there is (yet) a low number of studied samples which are inherent for ancient DNA studies, the results of Eindhoven do not reveal substantial frequency differences on the level of the main (sub-)haplogroups (Altena et al., 2013), which may suggest low influence of genetic drift and natural selection on the Y-chromosomal diversity within a community. Therefore, the genetic genealogical approach should be indeed a valuable alternative of ancient DNA to perform historical survey in West-European communities since surname adoption. Finally, future research with full genomic admixture analysis of regional sampling campaigns combined with appropriate statistical methods may provide more information about the influence of particular (un)known historical admixture events in West-European regions which are currently invisible by using the genetic genealogical approach based on the haploid markers (Hellenthal et al., 2014).

## DATA ARCHIVING

All genotypic data of this study have been archived in the open access Y-STR Haplotype Reference Database (YHRD, www.yhrd.org): accession numbers YA003651, YA003652, YA003653, YA003738, YA003739, YA003740, YA003741 and YA003742.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We thank all the volunteers who donated DNA samples. We acknowledge all the personnel who were involved in the collection of samples and genealogical data: Marc Van den Cloot, Valerie Vermassen, Werner Peene, Marcel Desmedt, Carlos Aelvoet, Lambert Joosten, William Dehenau and Annemie Swillens (Familiekunde Vlaanderen), Steven Vandewal (Stadsarchief Tongeren), Freddy De Chou (Gerardimontium), Jaak Heeren, Jos Grosemans and Maurice Smeers (Geschiedkundige Kring Alken), Marcel Cock and Marc De Ruyver. We also thank Luc van Durme, Guido Creemers, Danny Lamarcq, Wouter Dhaeze, Janne Swaegers, Brigitte Meijns, Luc De Meester, Tom Wenseleers, Marie Boz, Gilles Defraene, Michelle Panting and Lucrece Lernout for useful assistance and discussions. Finally, we thank three anonymous referees for their valuable comments on a previous version of this manuscript. ML is a postdoctoral fellow of the FWO-Vlaanderen (Research Foundation-Flanders). NB has a PhD-VLADOC scholarship of VLIR-UOS. This study was funded in the context of the 'Romeins DNA-project' by the Province of Limburg, FWO-Vlaanderen (Project number 1.5.116.13 N), the community of Jabbeke, the KU Leuven BOF-Centre of Excellence Financing on 'Eco- and socio-evolutionary dynamics' (Project number PF/2010/07) and on 'Centre for Archaeological Sciences 2-New methods for research in demography and interregional exchange' and the KU Leuven BOF-GOA 'New approaches to the social dynamics of long-term fertility change' (Project number 3H130264).

- Altena E, Smeding R, de Knijff P (2013). Het DNA-onderzoek: de eerste resultaten. In: Arts N editor. Een Knekelveld Maakt Geschiedenis. Utrecht: Uitgeverij Matrijs, p 213–227.
- Anderson KG (2006). How well does paternity confidence match actual paternity? Evidence from worldwide nonpaternity rates. *Curr Anthropol* 47: 513–520.
- Ballantyne KN, Goedbloed M, Fang RX, Schaap O, Lao O, Wollstein A et al. (2010). Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87: 341–353.
- Barrai I, Rodriguez-Larralde A, Manni F, Ruggiero V, Tartari D, Scapoli C *et al.* (2004). Isolation by language and distance in Belgium. *Ann Hum Genet* **68**: 1–16.
- Bechert T (2007). Germania inferior. Eine provinz an der nordgrenze des römischen reiches. *Mainz-am-Rhein: Philip van Zabern.*
- Boattini A, Pettener D (2013). What can we do with surnames today? An anthropologist point of view: Comment on "Surname distribution in population genetics and in statistical physics" by Paolo Rossi. *Phys Life Rev* **10**: 418–419.

- Bowden GR, Balaresque P, King TE, Hansen Z, Lee AC, Pergl-Wilson G et al. (2008). Excavating past population structures by surname-based sampling: the genetic legacy of the Vikings in northwest England. *Mol Biol Evol* 25: 301–309.
- Busby GBJ, Brisighelli F, Sánchez-Diz P, Ramos-Luis E, Martinez-Cadenas C, Thomas MG et al. (2012). The peopling of Europe and the cautionary tale of Y chromosome lineage R-M269. Proc Royal Soc B 279: 884–892.
- Cloet M, Vandenbroeke C (1989). Tien bijdragen tot de lokale en regionale demografie in Vlaanderen. Brussel: Gemeentekrediet, p 291.
- Cruciani F, Trombetta B, Antonelli C, Pascone R, Valesini G, Scalzi V *et al.* (2011). Strong intra- and inter-continental differentiation revealed by Y chromosome SNPs M269, U106 and U152. *Forensic Sci Int Genet* 5: E49–E52.
- Debrabandere F (2003). Woordenboek van de familienamen in België en Noord-Frankrijk. Amsterdam/Antwerpen: L.J. Veen/Het Taalfonds.
- DeWitte SN (2014). Mortality risk and survival in the aftermath of the Medieval Black Death. PLoS ONE 9: e96513.
- Excoffier L, Laval G, Schneider S (2005). ARLEQUIN ver.3.0: an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1: 47–50.
- Faure E, Royer-Carenzi M (2008). Is the European spatial distribution of the HIV-1-resistant CCR5-Delta 32 allele formed by a breakdown of the pathocenosis due to the historical Roman expansion? *Infect Genet Evol* 8: 864–874.
- Fincher CL, Thornhill R (2008). Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religion diversity. *Proc Biol Sci* 275: 2587–2594.
- Gotelli NJ, Colwell RK (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecol Lett* **4**: 379–391.
- Halsall G (2007). Barbarian Migrations and the Roman West. Cambridge: Cambridge University Press, p 376–568.
- Heck KL, van Belle G, Simberloff D (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* 56: 1459–1461.
- Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefansson K (2003). A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet* 72: 1370–1388.
- Hellenthal G, Busby GBJ, Band G, Wilson JF, Capelli C, Falush D et al. (2014). A genetic atlas of human admixture history. Science 343: 747–751.
- Hillewaert B, Hollevoet Y, Ryckaert M (2011). *Op het raakvlak van twee landschappen De vroegste geschiedenis van Brugge*. Brugge: Raakvlak Brugge.
- Hurlbert SH (1971). The nonconcept of species diversity: a critique and alternative parameters. *Ecology* **52**: 577–586.
- Jobling M, Hollox E, Hurles M, Kivisild T, Tyler-Smith C (2013). *Human Evolutionary Genetics*. New York, USA: Garland Science, p 650.
- Jobling MA (2012). The impact of recent events on human genetic diversity. *Philos Trans R Soc Lond B Biol Sci* 367: 793–799.
- Jones S (1997). The Archaeology of Ethnicity. Constructing Identities in the Past and Present. London/New York: Routledge.
- Kayser M, Ballantyne KN (2014). Y chromosome in forensic science. In: Primorac D, Shanfield M editors. *Forensic DNA Applications: An Interdisciplinary Perspective*. CRC Press Taylor & Francis Group.
- King TE, Jobling MA (2009a). Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Mol Biol Evol* 26: 1093–1102.
- King TE, Jobling MA (2009b). What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* 25: 351–360.
- King TE, Parkin EJ, Swinfield G, Cruciani F, Scozzari R, Rosa A et al. (2007). Africans in Yorkshire? The deepest-rooting clade of the Y phylogeny within an English genealogy. Eur J Hum Genet 15: 288–293.
- Lamarcq D, Rogge M (1996). *De Taalgrens. Van de oude tot de nieuwe Belgen.* Leuven: Davidsfonds.
- Lao O, Altena E, Becker C, Brauer S, Kraaijenbrink T, van Oven M et al. (2013). Clinal distribution of human genomic diversity across the Netherlands despite archaeological evidence for genetic discontinuities in Dutch population history. *Investig Genet* 4: 9.
- Larmuseau M, Vanoverbeke J, Van Geystelen A, Defraene G, Vanderheyden N, Matthys K et al. (2013a). Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data. Proc Biol Sci 280: 1772.
- Larmuseau MHD, Ottoni C, Raeymaekers JAM, Vanderheyden N, Larmuseau HFM, Decorte R *et al.* (2012a). Temporal differentiation across a West-European Y-chromosomal cline–genealogy as a tool in human population genetics. *Eur J Hum Genet* **20**: 434–440.
- Larmuseau MHD, Van Geystelen A, van Oven M, Decorte R (2013b). Genetic genealogy comes of age-perspectives on the use of deep-rooted pedigrees in human population genetics. *Am J Phys Anthropol* **150**: 505–511.
- Larmuseau MHD, Vanderheyden N, Jacobs M, Coomans M, Larno L, Decorte R et al. (2011). Micro-geographic distribution of Y-chromosomal variation in the central-western European region Brabant. Forensic Sci Int Genet 5: 95–99.
- Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R et al. (2014a). Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance. Ann Hum Genet 78: 92–103.
- Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, Kayser M, Decorte R et al. (2014b). Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations. *Forensic Sci Int Genet* **9**: 179–185.

- Larmuseau MHD, Vanoverbeke J, Gielis G, Vanderheyden N, Larmuseau HFM, Decorte R et al. (2012b). In the name of the migrant father–analysis of surname origin identifies historic admixture events undetectable from genealogical records. *Heredity* 109: 90–95.
- Larmuseau MHD, Vanoverbeke J, Van Geystelen A, Defraene G, Vanderheyden N, Matthijs K *et al.* (2013c). Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data. *Proc Biol Sci* **280**: 20132400.
- Loveluck C, Tys D (2006). Coastal societies, exchange and identity along the Channel and southern North Sea shores of Europe, AD 600-1000. *J Mari Arch* **1**: 140–169.
- Mantel N (1967). The detection of disease clustering and a generalised regression approach. *Cancer Res* **27**: 209–220.
- Moore LT, McEvoy B, Cape E, Simms K, Bradley DG (2006). A Y-chromosome signature of hegemony in Gaelic Ireland. Am J Hum Genet 78: 334–338.
- Oksanen J, Kindt R, Legendre P, O'Hara RB (2007). Vegan. Community Ecology Package version 1: 8–5.
- Peakall R, Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research - an update. *Bioinformatics* 28: 2537–2539.
- Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R et al. (2014). A global analysis of Ychromosomal haplotype diversity for 23 STR loci. Forensic Sci Int Genet 12: 12–23.
- Ralph P, Coop G (2013). The geography of recent genetic ancestry across Europe. *PLoS Biol* 11: e1001555.
  Redmonds G, King T, Hey D (2011). *Surnames, DNA, & Family History*. Oxford: Oxford
- Redmonds G, King T, Hey D (2011). Surnames, DNA, & Family History. Oxford: Oxford University Press.

- Rice WR (1989). Analyzing tables of statistical tests. Evolution 43: 223-225.
- Schürer K (2004). Surnames and the search for regions. *Local Population Studies* 72: 50–76.
- Scully M, King T, Brown SD (2013). Remediating Viking origins: genetic code as archival memory of the remote past. Sociology 47: 921–938.
- Solé-Morata N, Bertranpetit J, Comas D, Calafell F (2014). Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed. Ann Hum Genet 78: 253–254.
- The R Community (2011). Foundation for Statistical Computing R. R version 2.13.0.
- Walsh B (2001). Estimating the time to the most recent common ancestor for the Y chromosome or mitochondrial DNA for a pair of individuals. *Genetics* 158: 897–912.
- Willigan JD, Lynch KA (1982). Sources and methods of historical demography. New York: Academic Press, p 528.
- Wilson Sayres MA, Lohmueller KE, Nielsen R (2014). Natural selection reduced diversity on human Y chromosomes. *PLoS Genet* 10: e1004064.
- Winney B, Boumertit A, Day T, Davison D, Echeta C, Evseeva I *et al.* (2012). People of the British Isles: preliminary analysis of genotypes and surnames in a UK-control population. *Eur J Hum Genet* **20**: 203–210.
- Work TT, Jacobs JM, Spence JR, Volney WJ (2010). Increased green-tree retention required to maintain ground beetle biodiversity in boreal mixed wood forests. *Ecol Appl* 20: 741–751.
- Zerjal T, Xue YL, Bertorelle G, Wells RS, Bao WD, Zhu SL *et al.* (2003). The genetic legacy of the Mongols. *Am J Hum Genet* **72**: 717–721.

Supplementary Information accompanies this paper on Heredity website (http://www.nature.com/hdy)

12