The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research

Francesc Calafell[1] & Maarten H.D. Larmuseau[2,3]

[1]Institute of Evolutionary Biology (CSIC-UPF), Departament de Ciències Experimentals i de la Salut, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain

[2]KU Leuven, Forensic Biomedical Sciences, Department of Imaging & Pathology, Leuven, Belgium

[3]KU Leuven, Laboratory of Socioecology and Social Evolution, Department of Biology, Leuven, Belgium

Corresponding authors:

- Prof. dr. Francesc Calafell, Institute of Evolutionary Biology (CSIC-UPF), CEXS-UPF-PRBB, Doctor Aiguader 88, 08005 Barcelona, Catalonia, Spain. Phone: +34-93.316.08.42; fax: +34-93.316.09.01; e-mail: francesc.calafell@upf.edu

- dr. Maarten Larmuseau; KU Leuven – Catholic University of Leuven, Forensic Biomedical Sciences, Kapucijnenvoer 33, B–3000 Leuven, Belgium; Fax: +32 (0) 16324575, Email: maarten.larmuseau@bio.kuleuven.be

Running title: Y chromosome popular in genetic genealogy

Keywords: Genetic genealogy, Y chromosome, Interdisciplinary research, Extra-pair paternity, Surnames, Patrilineage

Abstract

The Y chromosome is currently by far the most popular marker in genetic genealogy that combines genetic data and family history. This popularity is based on its haploid character and its close association with the patrilineage and paternal inherited surname. Other markers have not been found (yet) to overrule this status due to the low sensitivity and precision of autosomal DNA for genetic genealogical applications, given the vagaries of recombination, and the lower capacities of mitochondrial DNA combined with an in general much lower interest in maternal lineages. The current knowledge about the Y chromosome and the availability of markers with divergent mutation rates make it possible to answer questions on relatedness

levels which differ in time depth; from the individual and familial level to the surnames, clan and population level. The use of the Y chromosome in genetic genealogy has led to applications in several well-established research disciplines; namely in e.g. family history, demography, anthropology, forensic sciences, population genetics and sex chromosome evolution. The information obtained from analysing this chromosome is not only interesting for academic scientists but also for the huge and lively community of amateur genealogists and citizen scientists, fascinated in analysing their own genealogy or surname. This popularity however has also some drawbacks, mainly for privacy reasons related to the DNA donor, his close family and far-related namesakes. In this review paper we argue why Y-chromosomal analysis and its genetic genealogical applications will still perform an important role in future interdisciplinary research.

Introduction

The quest for one's roots seems to be a universal endeavour. The desire to find one's origin is often translated into tracing back a genealogical tree and identifying one's ancestors (Zerubavel 2012). This is usually done by researching archival evidence, such as documents in civil or parish records that can reveal genealogical connections. These original genealogical records are increasingly being made available online through national and local archives, open access sites like http://familysearch.org, and several subscription sites of commercial genealogical companies. This type of evidence, however, is quite often lost or may have never existed in the first place. In that case, one can resort to genetic tests, since they can establish the genetic relatedness between individuals at a time depth comparable to that of most genealogical trees (Jobling et al. 2013). A key difference can be noticed: genetic tests establish only biological and not social relatedness. Adoption or false paternity may result in socially defined familial relations that are not genetic. Thus, genetic tests provide also the possibility to verify the biological *versus* legal genealogy of an individual, something which cannot be verified using only written records.

The whole genome can be (and is being) used to establish relatedness. At the single locus level, identity by state among alleles can be translated into identity by descent and hence into some relatedness index if the population allele frequencies are known (Speed and Balding 2015). If whole chromosomes are considered (which is feasible with SNP arrays or whole genome sequencing), shared segments can be determined. The number and length of these segments

can be used to determine the number of meioses that separates two individuals (Lareu et al. 2012). Beyond close family (>6th degree relationships), however, both methods suffer from low sensitivity and precision even when using the whole genome, due to the vagaries of recombination (Byrnes et al. 2014; Huff et al. 2011; Li et al. 2014; Speed and Balding 2015). Recently developed methods based on rare shared variants in whole genome data enlarge the possibilities to allow detecting relationships of a higher degree (Al-Khudhair et al. 2015; Schiffels et al. 2016), however, their sensitivity is still poorer than that afforded by haploid markers. Unilinearly transmitted genome segments, namely mitochondrial DNA (mtDNA) and the non-recombining portion of the Y chromosome (NRY) are strictly passed down through respectively the maternal and paternal lines, and could in theory be used to trace them. Note that, given the lack of recombination in these genome regions, they have been inherited through the genealogical tree that unites either all human maternal lineages or all paternal lineages (Underhill and Kivisild 2007).

In genetic genealogy, the NRY is clearly more popular than mtDNA for a number of reasons. The Y chromosome has a stronger geographic differentiation compared to mtDNA, probably because of a higher female migration rate linked to traditional patrilocal marriages (i.e., the bride moving into the groom's village) (Wilkins 2006) and this especially on a local scale (Marks et al. 2012). Also, as discussed in the next paragraph, the polymorphisms in the Y chromosome have a much wider range of mutation rates than those in mtDNA, which are practically limited to nucleotide substitutions and short indels (van Oven and Kayser 2008). Thus, the time range over which the Y chromosome provides reliable estimates of the time to the most recent common ancestor (tMRCA) spans from a few generations to thousands of years. Finally, most societies are patriarchal, and family relations are defined on the male line. A clear expression of this pattern are surnames: the Y chromosome traces surnames, while the surname of the bearer of a particular mtDNA changes each generation. Therefore, it is understandable that the paternal line is the most interesting and best studied among genealogists. Moreover, since many of the academic Y-chromosome experts and of the citizen scientists include woman researchers, the popularity of Y-chromosome is not related to a male-biased interest in genetic genealogy (Zerubavel 2012).

Diversity is required for Y chromosomes to be able to detect male relatedness. Two types of genetic polymorphisms are mostly used in Y-based genetic genealogy: single nucleotide polymorphisms (SNPs) and short tandem repeats (STRs). The NRY is nearly 57 Mb long (Quintana-Murci and Fellous 2001) and at least 65,000 SNPs and short indels have been found

by sequencing 1,244 human Y chromosomes from the 1000 Genomes Project (Poznik et al. 2016). Each one of these Y-SNPs first originated at some branch of the genealogical tree that unites all human males. Some Y-SNPs may have deep origins and are shared by most men, others may be more recent and are spread in some geographical regions, whereas others may be so recent that they are confined to a single family or even a single individual. Nevertheless, their sheer number has allowed to reconstruct the main branches of the paternal phylogenetic tree which are referred to as *haplogroups*. Population studies have mapped the geographical deployment of this tree, often referred to as the phylogeographical structure of the Y chromosome (Jobling and Tyler-Smith 2003). In genetic genealogy, Y-SNPs can thereby be used to establish the evolutionary lineage and the broad geographic origin of a particular Y chromosome. Y-STRs are less numerous than Y-SNPs: the most comprehensive catalogue lists 4500 of them (Willems et al. 2016). However, they have faster mutation rates than Y-SNPs: the average mutation rate for all polymorphic Y-STRs is 3.83×10^{-4} mutations per generation (Willems et al. 2016), although the most commonly employed and commercially available subsets have faster mutation rates: 2.80×10^{-3} (www.yhrd.org, accessed June 16th, 2016) for the 16 most used Y-STRs, and 13 rapid-mutating Y-STRs reach mutation rates $>10^{-2}$ (Ballantyne et al. 2012). Combination of Y-STR alleles are referred to as *haplotypes*. Since each Y-STR mutation happens in some Y chromosome with a particular Y-SNP haplogroup, Y-STR allele variation is deeply partitioned by haplogroup (Bosch et al. 1999) after which Y-STR haplotypes can be used with varying precision to predict Y-chromosomal haplogroups in the absence of Y-SNP genotypes (Athey 2005; Athey 2006; Schlecht et al. 2008). The occurrence of duplicated Y-STR alleles and intermediate Y-STR allele variants is especially useful in haplogroup predicting as they are often linked to a specific haplogroup (Balaresque et al. 2009; Myres et al. 2007).

Both Y-SNPs and Y-STRs can be used to detect whether two men are patrilineally related. However, only subsets of Y-SNPs that are used to determine haplogroups are of course not sufficient to differentiate each patrilineal family within a population: after typing 120 SNPs, Larmuseau *et al.* (2014c) found that 6.7% of the pairs among 773 unrelated Flemish men carried identical haplogroups. But, using comprehensive Y-chromosome sequencing tests - a service that some companies also offer to the general public, namely Full Genomics Corporation (Rockville, Maryland, USA) and Family Tree DNA (Houston, Texas, USA) - relatedness can be easily ascertained including the tMRCA, provided that sequencing has been performed to the depth, required to call appropriately all Y-SNP genotypes. For instance, the closest pair of Y chromosomes in a sample of 50 Spanish males (Poznik et al. 2016) has different alleles in 29 SNPs (out of 5,562 polymorphic positions in 12,923,389 bp of sequence),

while with a pedigree mutation rate of 3.07×10^{-8} per basepair and generation (Helgason et al. 2015), the probability of having more than 10 differences in a pair of fourth cousins is 0.27% (estimated under an infinite sites model with a Poisson distribution). Closer relationships would obviously bear even more similar Y chromosomes. As for Y-STRs, one of the most popular sets, i.e., the *PowerPlex® Y23* System (Promega, Madison WI, USA), yields a closer call: 0.006% of the pairs among 705 unrelated Spanish males (Purps et al. 2014) carry identical 23-Y-STR haplotypes, while we estimate that 39.83% of fourth cousin pairs would have identical STR haplotypes when using locus-specific mutation rates (www.yhrd.org, accessed June 16th, 2016) and a strictly stepwise mutation model (Figure 1). A more definite separation could be obtained by adding rapidly-mutating Y-STRs (Ballantyne et al. 2014) or Y-SNPs within the R1b haplogroup (Larmuseau et al. 2014b; Solé-Morata et al. 2014). In general, the combination of appropriately selected Y-SNPs and large numbers of Y-STRs makes it possible to fine-tune genetic genealogic testing (Walsh 2001).

In the following sections, we will discuss the application of the Y chromosome in a genealogic frame to identify historically relevant remains, to confirm genealogical relatedness (or conversely, to estimate false-paternity rates), and to ascertain patterns of relatedness in communities and populations. We will observe how the Y chromosome can be used to learn about a system of memes that are paternally inherited as well, namely surnames. Beyond genealogy, we will also venture into the current and future applications of Y chromosome lineages, in fields going from demography to forensic genetics. Finally, we will describe how genetic genealogy is one of the scientific pursuits where the work of citizen scientists has had a sizeable contribution, but we will also stress how its popularity can lead to ethical concerns.

Genealogy, identification, and the Y chromosome

Human remains can be identified by genetic comparison with undisputed samples from the same individual or with samples from close relatives. However, if the relatives being compared are a few meioses apart, the shared fraction of the autosomal genome decreases exponentially and autosomal markers lose much of their power to produce a positive identification. The Y chromosome does not experience such decay in power and the differences between relatives accrue only as fast as the mutation rate of the polymorphisms being compared (Kayser and Ballantyne 2014). Coupled to the fact that male-line relatives are easier to trace through their surnames, this implies that the Y chromosome has been used to identify historically relevant

individuals. In some cases, the identification of the dead in violent conflicts is sought. For instance, human remains in mass graves from World War II (Pajnic et al. 2010) and the Spanish Civil War (Baeta et al. 2015) were identified with Y-STRs proving to be an invaluable tool, alongside autosomal STRs. Other cases were well beyond the capabilities of autosomal STRs; thus, Y-SNPs and Y-STRs have been used to identify a number of remains of historical figures. A good example is the research on the remains of Jörg Jenatsch, who was a Swiss national hero in the 17th century. His putative remains were exhumed in 2012 from the Chur cathedral. Three Jenatsch men, separated by 14 meioses from Jörg Jenatsch, carried the same Y haplogroup as found in the Chur remains and differed only in one repeat each at three Y-STRs. According to the calculations by Haas *et al.* (2013), this makes the remains 20 times more likely to belong to Jörg Jenatsch than to an unknown male.

After this success, we discuss two other cases that provide cautionary tales, rather than positive identifications using Y-chromosomal data. In 2012, a skeleton was excavated at the presumed site of the Grey Friars friary in Leicester, the last-known resting place of King Richard III. Archaeological, osteological and radiocarbon dating data were consistent with being his remains; mtDNA comparison with two living maternal relatives (although separated by 18 and 20 meioses) gave a match. Taking all this in consideration, this was all positive evidence for the identification of King Richard III (King et al. 2014). But when a set of five paternal relatives were compared to each other, four shared a Y-STR haplotype, while the fifth did not. The shared Y haplogroup did not match with that of the King's remnants. Given the overwhelming evidence for the remains being those of Richard III, King *et al.* (2014) uncovered two historical cases of false paternity, one in the four generations in the branch leading to the living relative carrying a different Y-STR haplotype and one in the 18 generations intervening between the common ancestor of the five living relatives and Richard III. Although male-line relatives are easier to find from genealogical records and peerage registries, false paternity is a much more common occurrence than false maternity, and has to be taken into account in identification studies based on Y-chromosomal data.

Other problems may arise from the difficulties in extracting sufficient undegraded DNA from ancient samples, and from the fact that the remains of historical figures can become coveted relics with monetary value and prey to forgery. Both factors may have been in play in the attempt to verify the presumed remains of Bourbon kings of France: a mummified head attributed to Henry IV, and a handkerchief that may have been dipped in Louis XVI's blood. Seven generations separated the two kings on the male line, and a six Y-STR haplotype showed

only one difference in repeat among the two relics (Charlier et al. 2013; Lalueza-Fox et al. 2011). However, an analysis of three living male relations, belonging to two different lines that had common Bourbon ancestors which were direct descendants of Henry IV and ancestors of Louis XVI, showed that all three shared a haplogroup and had similar 38 Y-STR haplotypes, which were different from those in the relics (Larmuseau et al. 2014a). Moreover, Henry IV was maternally related to Louis XVII, and a mtDNA sequence recovered from his preserved heart did not match that from the mummified Henry IV's head (Jehaes et al. 1998; Jehaes et al. 2001). Later, the authors of the original paper on the Louis XVI's relic went on to sequence the entire genome of the blood on the handkerchief (Olalde et al. 2014). The results placed the ancestry of the relic in northern Italy (coincidentally, where the relic surfaced), while Louis XVI's ancestry was mostly central European. Moreover, some phenotypical traits that could be predicted from the genomic sequence (such as height, obesity and eye colour) did not match Louis XVI's traits. The authors argued that either the relic was forged or it was heavily contaminated with other human sources, and in any case, the partial match with Henry IV's relic was sheer coincidence.

Confirming genealogical relatedness

Document-based genealogy is the first resource to be employed in searching family roots or in establishing whether two people are related. However, the paper trail may be missing, or it may not reflect the biological relationship between two people. We have already seen how the Y chromosome is a powerful tool to discern a genetic connection by analysing Y-SNPs and Y-STRs, although it is not particularly precise to establish the exact degree of relatedness. This power has been used to solve some notorious cases involving public figures. The most famous case is that of US President Thomas Jefferson: by comparing his respective male-line descendants, it was confirmed that Thomas Jefferson – and not one of Jefferson's sister's sons as was initially believed - fathered his slave Sally Heming's last son (Foster et al. 1998). Also, it was refuted that Chinese emperor Cao Cao (155AD-220AD) was related to marquis Cao Can, as he claimed (Wang et al. 2012) by comparing their respective descendants. Subsequently, the ancient DNA analysis of emperor Cao Cao's granduncle confirmed his relatedness to his putative current descendants (Wang et al. 2013).

Rather than seeking genetic confirmation of genealogical relationships, Y chromosome analysis has been used to quantify rates of misattributed paternity or non-paternity (that is,

discrepancy between the social and the biological father) by comparing large numbers of distantly related men in populations where oral or documental genealogies can be obtained. Note that this method would not detect non-paternities produced by patrilineal relatives of the social father. Strassman *et al.* (2012) found that the rate of non-paternity events was 1.8% per generation in Dogon (Mali) across all religions. It should be taken into account that this study was based on oral genealogies, which could lead to some errors and an overestimation of non-paternity rates. Studies based on written genealogies have produced similarly low rates: 0.9% per generation in Flanders, Belgium across the last 500 years (Larmuseau et al. 2013b), which was exactly the same estimated rate for the last 300 years in the Afrikaner population in South Africa (Greeff and Erasmus 2015) but slightly lower than that in a North Italian population (1.2% per generation over the last 400 years) (Boattini et al. 2015).

In the name of the father

Surnames and the Y chromosome are inherited together through the paternal line in most cultures with exceptions due to false paternity, adoption and inheritance of the maternal surname. Actually, surnames can be modelled as alleles of a single locus in the Y chromosome and have been used as such to investigate inbreeding and gene flow between populations through measurement of isonymy (that is, sharing surnames) (Colantonio et al. 2003). Currently, the reverse approach is more common: the genetic diversity can be used to investigate surnames, whether piecemeal in particular cases, or the overall surname system of a population. In the first scenario, the relatedness of the individuals carrying a surname can be easily determined by Y-SNPs and Y-STRs, and thus the common origin of that surname can be established. Some companies provide the public with the genetic tools necessary to investigate their own surnames, and as we will discuss below, single-surname investigation is mostly the province of citizen scientists. However, a few such studies in surnames of some interest have been undertaken in academia. Sykes and Irven (2000) demonstrated for the first time that indeed different groups of men carrying the Sykes surname had different Y chromosome genetic profiles, and thus were descendants of different founders of the surname. In the same year an identical study was carried out in Chinese men bearing the Wang surname but no differences in Y-chromosomal variants were found with a random Chinese sample, which was explained by the complex origins of the surname (Dongying et al. 2000). Other surnames have been investigated by their historical relevance. Martínez-González *et al.* (2012) showed that the Y chromosomes of bearers of the Colombo and Colom surnames

(which would have been the original surname of Christopher Columbus if he were of Genovese or Catalan descent, respectively) were different from each other; moreover, while Colom Y chromosomes formed clear distinct subsets that mapped to different geographical origins within eastern Spain, the Colombos harboured a huge genetic diversity with most Y haplotypes being represented just once. This pattern matches the history of the surname, which was often given to orphans and foundlings in Milan. The authors concluded that, if DNA from Columbus' remains could ever be extracted and genotyped, a Catalan origin could easily be proven, while a match in Italy would be much harder to find. Finally, Martinez-Cruz *et al.* (2012) showed that carriers of the Basarab surname were not related to the Basarab dynasty of Vlad III The Impaler, also known as Dracula, since the diversity in their Y chromosomes was much shallower and originated in the 19th century rather than the 15th century when Vlad III ruled, or than the 11th century, when the dynasty originated.

Studies on whole surname systems of populations are more interesting than particular case studies. King *et al.* (2006) compared pairs of unrelated British men carrying the same surname and found that signals of co-ancestry were stronger in rarer surnames. They also investigated the intriguing possibility of predicting the surname of the unknown donor of a biological sample, such as in a forensic investigation. King and Jobling (2009a) analyzed British surnames and estimated haplotype diversity and the number of founders. The results showed that in Britain, the more frequent surnames tended to have more diverse Y-chromosomal haplotypes, as if surname frequency was driven by polyphyletism. In contrast, an analog study carried out in Ireland (McEvoy and Bradley 2006) showed no significant correlation between the frequency of a surname and the diversity of the Y chromosomes within it. Moreover, the finding that frequent Irish surnames such as O'Sullivan and Ryan present a single founder may be accounted for by male social and reproductive success in the past. Consequently, they considered that the Y chromosome could have suffered from the effect of natural selection not due to its genetic content but to a cultural marker: the surname. On the other hand, it should be noted that this study used only 17 Y-STRs which is insufficient to rule out coincidental matching as a result of convergence, which is especially the case within the – in Ireland most frequent – haplogroup R1b (Larmuseau et al. 2014b; Solé-Morata et al. 2014). Therefore, a subsequent high-resolution study in Ireland is required to verify the claims by McEvoy and Bradley (2006). Nonetheless, in Spain, either using Catalan surnames (Solé-Morata et al. 2015) or Spanish, Catalan and Basque surnames (Martinez-Cadenas et al. 2016), the British model is followed in which the frequency of a surname is driven by how often it was founded. They also quantified the rate of introgression of new Y chromosomes into a surname at 2-3% per

generation, and found that the higher rates in areas where the maternal surname was transmitted (if a woman inherited an estate, she transmitted her surname, as a way to ensure that the connection between the land and the surname survived). Also in Andalusia (Spain) Calderón *et al.* (2015) failed to find such a relationship and interpreted this as a result of the late acquisition of Christian surnames by the local Moslems and Jews. However, they sampled regardless of surname, with many rare (and presumably monophyletic) surnames being sampled just once.

Clans, lineages, castes and other patrilineal institutions

Next to the patrilineal surname – which has a quite recent historical development - there are many other types of descent groups in human populations whereby membership is patrilineally inherited (Kottak 2002). Members of these groups claim to have a common paternal ancestor or a small group of founder ancestors. Y-chromosomal analysis has the possibility to test the biological relatedness among members of such patrilineal descent groups. The presence of one main descent cluster or a limited number of clusters, and/or a reduction of Y-chromosomal variation in the tested group *versus* the global population, is generally recognized as a biological basis for these social structures. In the past decade, scientists already studied Y chromosomes within and between patrilineal descent groups in traditional societies as the so-called 'lineages', 'clans' and 'tribes' which are based on oral genealogies and narrative claims (Chaix et al. 2004; Sanchez-Faddeev et al. 2013). But it is not only limited to traditional societies since it is also present and already analysed in Western and Asian populations, like clans in Ireland (McEvoy et al. 2008; Moore et al. 2006), the 1800-year old stipulated ancestries in China (Wang et al. 2012; Wang et al. 2013), the patrilineal Jewish priest caste (Skorecki et al. 1997) and the privileged institution *Partecipanza* in Northern Italy (Boattini et al. 2015). Therefore, this type of analysis is complementary to surname studies. However, two generalities are notable in this research. First, an older (estimated) origin of the descent group will lead to a lower signal with Y-chromosomal variation in comparison with the global population. This is partly due to the effect of telescoping, which occurs when there is no fully demonstrated or stipulated descent anymore and the genealogy lacks a gap between the recent past and the (sometimes mythical) 'time of origin' (Chaix et al. 2004; Sanchez-Faddeev et al. 2013). Contrariwise, there is also a cumulative effect of adoption and extra-pair paternity disturbing the association between the Y chromosome and patrilineal membership since the origin of the descent group. Second, although the majority of the members of a specific

descent group has a specific Y-STR haplotype and belongs to a phylogenetically in-depth Y-SNP haplogroup, not all of the individuals who are assigned to that specific Y-chromosomal variant have the same common recent paternal ancestor as those of the members of the particular patrilineal group. This leads to limitations of the use of Y-chromosomal variants as genealogical marker, which was already demonstrated with the so-called Cohanim haplotype (Skorecki et al. 1997) – including the extended Cohanim haplotype (Hammer et al. 2009) – which is associated with the Jewish priesthood but was already distributed across the Middle East before the origin of the Jewish priest tradition (Tofanelli et al. 2014).

Patrilineal relatedness without* a priori *expectation

Even if there is no expectation of patrilineal relatedness, Y-chromosomal analysis can still detect males sharing a most recent common ancestor who lived in genealogical or historical times. One intriguing phenomenon is the presence of large descent groups detectable on population level with a common ancestor who lived in a time period which is still relevant for genealogists and historians. These cases are assumed to be the result of strong 'social selection' in the past, as genetic drift cannot be the (only) driver for the wide distribution of that particular Y-chromosomal cluster and as natural selection on the few Y-chromosomal genes is assumed to be almost non-existent (Balaresque et al. 2015; Zerjal et al. 2003). Thus strong social selection with a high reproductive success of a certain male and his descendants is assumed to be the reason for these bursts in the genetic landscape. Several of those large descent groups with origins in the historical period are already described in Asian populations (Balaresque et al. 2015), including the ones attributed to descendants of the legendary Mongolian leader Genghis Khan (c. 1162-1227) (Zerjal et al. 2003) and of Giocangga (died in 1582) linked to the Qing dynasty (1644-1912) (Xue et al. 2005). The attribution of the large descent groups to both historical figures are based on the distribution and time estimations of the origin of high successful Y-STR lineages. Moreover, men whose Y chromosomes match these defined lineages, certainly descend from two prolific male ancestors which were carriers of the widespread Y-lineages and which had descendants who had a high reproductive success as well as wealth and socio-political power. These profiles are appropriate for Genghis Khan and Giocangga but the identities of these ancestors remain unclear without the analysis of ancient DNA from the attested remains of candidates or a long-dead descendant as ultimate proof. Next to the eleven examples of successful Y-lineages in the Asian population detected by Balaresque *et al.* (2015), another successful lineage was found in Ireland. Almost 20% of the

northwest Irish males belong to a Y-chromosomal lineage attributed to the early medieval 'Uí Néill' dynasty and the warlord Niall of the Nine Hostages who lived in fifth- or sixth-century Ireland (Moore et al. 2006). Although the remains of this legendary king or one of his descendants in the 'Uí Néill' dynasty have not been directly genotyped, the interpretation of the results is supported by the over-representation of the lineage in a large group of Irish surnames assumed to originate in the Uí Néill dynasty. On the other hand, historians challenged the assignment of this wide-spread Y-chromosomal lineage in Ireland to the early medieval 'Uí Néill' dynasty due to the lack of any link between the dynastic families recorded on the fifth-century texts and the later surnames of the tenth-twelfth centuries (Swift 2013). Moreover, the research by Moore et al. (2006) should be replicated on a higher Y-chromosomal resolution as the conclusions were drawn from haplotypes based on just 17 Y-STRs without the corresponding detailed Y-SNP testing to rule out coincidental matching as a result of convergence (Larmuseau et al. 2014b; Solé-Morata et al. 2014).

Outside these few examples of large descent groups on population level, there is limited knowledge about the chance of finding two males whose Y chromosomes match due to a shared genealogical or historical paternal ancestor, without any *a priori* expectation of relatedness as having the same (or similar) surname or clan membership. One study investigated the relatedness between Y chromosomes of different surnames within and between several communities in Flanders on the moment of the start of the surname adoption between the 14th and 15th century (Larmuseau et al. 2015). A low chance (0.35 ± 0.16%) of relatedness in historical and genealogical times between two males with a different indigenous surname was found based on in-depth Y-chromosomal genotyping, which was not higher than the chance of relatedness of donors assigned to the same region or assigned to entire Flanders. This was surprising as synchronic analyses for each time point in the last 400 years showed strong patrilocality in those small Flemish communities. This low chance of relatedness is surely the result of the high genetic drift – or the effect of daughtering out as known by genealogists – next to a high migration rate during the late Middle Ages in Western Europe at the time of introduction of surnames, a phenomenon which was already observed at the same time in England due to the Black Death (Redmonds et al. 2011). Next, several claims of unexpected patrilineal relatedness between males have been also yet made in on-line genealogical forums by citizen scientists but mostly always reported without any clear time estimate of the most recent common ancestor or statistical test to verify the chances for a false positive relatedness. The most mediatised but academic unpublished claims are those matches between Y chromosomes of a skeleton and a local inhabitant where the bones were

unearthed, e.g. as already several times attempted in the Netherlands to enlarge the public interest for a local archaeological project (de Rooij and M'Charek 2015; Lauwerier 2012). Nevertheless, all these claims cannot provide further insights in the occurrence of common patrilineal ancestry within a specific population or across a particular region, which is already attempted based on genomic data (Ralph and Coop 2013) and modelling (Rohde et al. 2004). Therefore, more population-wide studies using genetic genealogical data with accurate time estimates of Y-chromosomal matches are required.

Applications of the genealogical analysis of the Y chromosome

Y-chromosomal data linked to genealogies yield insights and provide applications in several research disciplines. In this review, we focus on specific applications in particular disciplines, although we are aware of the fact that these applications are never limited to a single discipline but are informative in several other disciplines since genetic genealogy is a catalyst in interdisciplinary research.

Family history

Obviously, genealogical applications of Y-chromosomal analysis are useful for anyone who is interested in family history. Genealogy is a popular hobby worldwide and is therefore often reported as the second most popular search term in online search engines (Rodriguez 2014). Today, genealogists are often turning to records written in DNA next to those in paper archives to solve their problems. As genealogists are generally interested in their paternal lineages, due to the fact that this is often coupled with the surname, Y-chromosomal tests genotyping Y-SNPs and Y-STRs are hugely popular with family historians (Zerubavel 2012). This enthusiasm was already experienced in the early days of genetic genealogy, even when the knowledge of Y-chromosomal variation and differentiation between far-related males was unable to provide accurate answers on many genealogical questions at that time (Brown 2002). In 2016, in-depth Y-chromosomal genotyping guarantees low type I and type II errors to verify patrilineal relatedness between two males as well as estimations of their time of most recent common ancestor (tMRCA) (Hughes and Page 2016). Therefore, several applications of Y-chromosomal analysis in family history were created. First, it is important to verify a particular patrilineal genealogy based on archival documents. Each step back in time requires making a decision by

linking at least two instances of information which can be complicated by spelling variation, variants (translations, suffix variation, etc.) and errors (Bloothooft and Schraagen 2015). Therefore, each genealogy is based on a number of decisions and is in fact always uncertain. Established relatedness based on Y chromosomes between two far-related DNA donors can provide additional evidence for the correctness of the legal genealogy (Fig.2a). Also specific disputable (recent or historical) steps in the genealogy can be verified to a certain level as Y-chromosomal variation is similar among patrilineally related males.

Second, when a 'paper-trail' seems irrefutable but the DNA evidence is contradictory, the genetic test can provide information about any form of illegitimacy somewhere in the genealogy which can be further explored by including extra DNA donors in the analysis (Fig.2a). Therefore, the Y chromosome provides a test for the comparison of legal and biological genealogy of a certain person (Greeff et al. 2012). Third, the most popular applications of Y-chromosomal analysis in family history are without any doubt the surname projects (Jobling 2001; King and Jobling 2009b). Not everyone with the same or a similar surname can be linked to each other in an established genealogy, and therefore (yet) unlinked bearers of a specific surname are puzzled if their surname is once independently given to different families or if they have a common paternal ancestor who cannot be traced back due to a lack of archival records. Therefore, a surname project can lead to the refinement of family trees and the inclusion of branches. Nevertheless, to be certain about the historical relatedness between two lineages with the same (or a similar) surname, participants always need to avoid as much extra-pair paternity and adoptions in the tested lineages as possible (Fig. 2b).

Finally, genealogists often make their Y-chromosomal profiles public or available in a match database (which may be general or restricted to a particular geographical region or haplogroup) in the hope to find an unexpected patrilineal relatedness. This may be one of the few possibilities for genealogists who have no clue about a paternal ancestor when their paternal lineages stops with a child who was not legitimised by a father, a foundling, an adoption or a sperm donor child. For others can an unexpected match reveal historical relatedness with another family before the adoption of surnames. Although these results may have some relevance for the families involved, there is only rarely academic interest in such case studies and therefore these applications are often called 'recreational genetics' (King and Jobling 2009b).

Demography and Anthropology

One central application of Y chromosomes in demography and anthropology is the possibility to estimate and compare (historical) rates of extra-pair paternity (EPP) within and between human populations. By verifying the relatedness within a large number of genealogies using Y-chromosomal data, an estimation of the historical EPP within a particular population can be realised (Larmuseau et al. 2016b). Therefore, a spatio-temporal analysis of EPP rates makes it possible to analyse in detail the influence of specific factors on paternity uncertainty; as e.g. socio-economic situation, religion (Strassmann et al. 2012), differences in paternal investments (Marlowe 2000; Sear 2016), urbanisation and industrialisation (Laslett et al. 1980), life history patterns and the introduction of reliable contraceptives (Harris 2016; Larmuseau et al. 2016c). This research creates therefore new opportunities to study courtship, marriages and families in the past (Gillis 1985; Rabb and Rotberg 1971) and to compare EPP rates with illegitimacy rates from well-established historic-demographical studies (Laslett et al. 1980). Moreover, Y-chromosomal research has the possibility to finally verify whether premarital children in a certain region and time period were later legitimised by their biological father or by another unrelated person; a fundamental question which is still unresolved in historical demography. Additionally, it is important to know the historical rate of extra-pair paternity within a population to verify the bias in evolutionary demographic studies and the analysis of biological traits, such as e.g. quantitative genetic estimates or the estimation of lifetime breeding success, using historical data (Bolund et al. 2015; Hayward et al. 2015).

Next to illegitimacy, as family structures and patrilineal social units like lineages, clans and tribes are highly relevant in many human societies, testing the biological background of those structures and institutions is as well highly informative for many anthropologists (Kottak 2002). By using Y-chromosomal markers, the veracity of oral genealogies and semimythical origins can be supported. Especially the difference in those claims between different types of societies, e.g. pastoral versus farmer populations (Chaix et al. 2007), is highly interesting to understand the evolution of social structures and to explain patrilineal cooperation (Strassmann and Kurapati 2016).

Forensic sciences

The analysis of the human Y chromosome provides a powerful tool in forensic cases for male sex identification, male lineage identification and identification of the geographical origin of male

lineages (Kayser and Ballantyne 2014). Especially in sexual assault cases, the Y chromosome provides essential evidence as it is often the only detectable trace of DNA of the offender (Roewer et al. 2014). Moreover, because male relatives share an identical Y chromosome profile for several generations, this chromosome is crucial in familial DNA searching or in the prediction of the surname within a forensic identification case (King and Jobling 2009b). Next, the combination of Y-chromosomal data and genealogical information can be applied in different ways in forensic research. First, the different available and future Y-STR kits can be tested to verify which of them accurately differentiate individuals between and within paternal lineages. Moreover, they can be used to improve familial DNA searching and surname prediction (Butler 2015). It is assumed that there has to be different Y-STR kits for different forensic questions, e.g. the 23 Y-STR kit is better for familial searching (Purps et al. 2014) in contrast to the rapid mutating (RM) Y-STRs which are better to differentiate males from one single paternal lineage (Ballantyne et al. 2014). Second, a database with surnames associated with Y-chromosomal genotypes, can be useful within forensic cases with no matches in a DNA database. As such, it is still possible to predict the surname of a particular victim or perpetrator. Nevertheless, as the link between surnames and Y-chromosomal data is weak for common names and not practical for rare names, it is assumed that a database with intermediate frequency surnames will be most efficient (King and Jobling 2009a; King and Jobling 2009b). Third, genetic genealogy can provide population genetic data of Y-chromosomal markers based on genealogically unrelated males. It is important that this data is also generated on different geographical levels, especially on communal scale in rural regions as the chance of uninformative haplotype matches in a particular region has to be known before a DNA survey is organised in the close area of a crime scene (Larmuseau et al. 2015). Fourth, genetic genealogical data provides accurate estimation of mutation rates of (new) Y-chromosomal markers based on (far-)related males. These analyses still increase the accuracy to estimate the tMRCA between two individuals (Ballantyne et al. 2010). Finally, specific forensic projects create a huge genetic genealogical database as one of the few possibilities to identify bodies in mass graves older than 50 years. By finding Y-chromosomal matches between skeletons and far-related individuals of potential victims, identification of long-dead individuals can still be realised like in the Fromelles project related to Australian and British soldiers in World War I (Scully 2014) and the project to identify victims of communists and Nazi totalitarian regimes in Poland (Ossowski et al. 2016).

Population genetics

The link between the Y chromosome, surname and paternal lineage provides an excellent opportunity to detect signals of (past) population stratification and migrations which are (yet) undetectable within genomic analysis of currently living individuals or based on the still very limited number of ancient DNA profiles (Larmuseau et al. 2013a). Since the genealogical data is linked to Y-SNPs which do not mutate in a genealogical time period, such genetic genealogical link makes it possible to study indirectly the population differentiation of each time period between the introduction of the patrilineal surnames until today (Larmuseau et al. 2012a). Therefore, it is possible to estimate the relevance of past demographic events as migrations during e.g. the Industrial Revolution in Western Europe at the beginning of the 19th century, the religious wars in the Low Countries at the end of the 16th century, both World Wars in the 20th century, *etc*. This approach also provides a more objective manner to explain an observed population genetic pattern than just link a particular observed pattern with a well-known past event by 'historical cherry-picking' (Jobling 2012). Of course, this genetic genealogical approach may only generate a virtual temporal sampling limited to individuals who had progeny until today and such a sample will therefore not necessarily represent the whole population at a certain point in the past (Helgason et al. 2003). Another approach is using surnames combined to in-depth genealogies in the sampling in two different ways. First, by analysing Y chromosomes of those males with surnames which were already present in the population since the introduction of surnames in the Late Middle Ages and which were not adopted during the last centuries. As such, it is possible to observe specific signatures of past gene flow events which are invisible in a random set of modern individuals currently living in a specific region or community (Boattini et al. 2015; Bowden et al. 2008; Larmuseau et al. 2015). Second, historical genetic admixture events that are undetectable from genealogical records can be identified by comparing specific groups of surnames classified by linguistic analysis (Larmuseau et al. 2012b; Niederstätter et al. 2012; Solé-Morata et al. 2015). Therefore, analyses of Y chromosomes linked to surnames and genealogies, together with a close interdisciplinary cooperation of geneticists, historians, linguists and archaeologists provide more insight in the dynamics of (intra-)regional migrations since the time of surname establishment till nowadays.

Sex chromosome evolution

The mechanisms driving sex chromosome evolution are currently a major research topic in evolutionary genomics (Hughes and Page 2015). Genetic genealogy has the possibility to study

differences of Y chromosomes within patrilineal lineages across 10-20 meioses. Although this is a relatively short time, analysing Y-chromosomal variation across deep-rooting pedigrees has already been highly useful in the detection and evolutionary analysis of an enlarged pseudo-autosomal region (ePAR) due to a non-allelic homologous recombination in particular Y-chromosomal lineages (Mensah et al. 2014), in the selection of relevant Y-chromosomal markers in the global (Van Geystelen et al. 2014) and minimal reference Y-chromosomal phylogeny (van Oven et al. 2014). It will also be interesting to further analyse the occurrence and mutation rate of copy-number variants (CNVs) on the human Y chromosome (Jobling 2008; Johansson et al. 2015) as well as the occurrence of X to Y gene-conversion events (Trombetta et al. 2010). More insights into the evolution of the human sex chromosomes and Y chromosome variance will definitely be provided by including full sequences of genealogical related persons in these analyses. Moreover, the analyses will also improve substantially the accuracy of genealogical and forensic applications using Y-chromosomal variation.

Huge popularity and drawbacks

Next to academic scientists, a huge and lively community of laypersons or so-called citizen-scientists are trying to answer genealogical questions by analysing Y-chromosomal data. Most popular are the surname projects and projects related to a specific geographic region to find unexpected far-related families. Amateurs have plenty of books, forums, Facebook groups, (video)conferences and websites at their disposal. The website of the International Society of Genetic Genealogy http://isogg.org provides educational resources and testing company comparison charts. Enthusiastic genealogists have even launched their own journal, namely the 'Journal of Genetic Genealogy', to publish case studies and articles outlining the methodology used by family historians (Pomery 2010a; Pomery 2010b). They also publish their genetic studies on particular surnames in e.g. the 'Journal of One-Name Studies' and in 'Nomina', the journal of the Society for Name Studies in Britain and Ireland (e.g. Meates 2006; Plant 2005; Plant 2007). Commercial companies offering direct-to-consumer (DTC) genetic genealogy testing for sale are therefore highly popular. Since the introduction of such DTC tests in 2000, genetic genealogy is an economic growth sector with continuously new players, mergers and acquisitions of companies (see an overview of the main companies over time: Brown 2002; King and Jobling 2009b; Phillips 2016; Royal et al. 2010; Shriver and Kittles 2004; Wagner et al. 2012; Wallace et al. 2015). The interest of citizen scientists is often helpful for academic scientists, especially in initiating research after the observation of unusual Y-

chromosomal variants (Mendez et al. 2013), providing enough DNA-donors belonging to a specific rare haplogroup (Sims et al. 2009), suggesting interesting Y-chromosomal markers (Rocca et al. 2012; www.isogg.org) and providing plenty of in-depth Y-chromosomal genotypes – comprehensive Y chromosome sequences at high coverage (50x-80x) and/or 111 Y-STRs – which are often not generated by academic studies due to budgetary considerations. Moreover, some citizen scientists also make sophisticated analyses and tools available, particularly with reference to the analysis of BAM files from Y-chromosome sequence data (e.g. http://www.ytree.net).

Although the DTC genetic genealogical tests themselves are reliable there are also drawbacks to its popularity, namely in the interpretation of data in relation to deep ancestry inferences and in several ethical issues associated with public access to Y-chromosomal data linked to genealogies. In popular science and on the internet Y-chromosomal markers have become widely identified with particular ancestral, historical or still existing groups such as Celts, Vikings, Anglo-Saxon or Jews. These claims are most likely based on population genetic analyses or distribution maps but they are wrongly applied on the level of an individual driven by public and commercial interests. As such the practice of individual genetic ancestry testing is often highly unreliable, as covered in detail in a recent publication by Jobling *et al.* (2016). Some of these claims were based on academic studies by geneticists using the *interpretative phylogeographic* approach which is now generally considered to be flawed based on simulations (Balloux 2010; Chikhi 2009; Nielsen and Beaumont 2009) and on large-scale ancient DNA analysis (Pickrell and Reich 2014). Often genetic genealogists are however still easily caught for these wrong ancestry claims, which was already notable from the start of genetic genealogical testing (Brown 2002). Although several initiatives of academic researchers were undertaken to counteract this so-called '*genetic astrology'* (Thomas 2013; http://www.ucl.ac.uk/mace-lab/debunking), it will still be very difficult in the future to refute such popular but unfounded deep ancestry propositions about specific Y-chromosomal variants.

Next to the difficulty to discern the border between scientific evidence and myth, the popularity of the Y chromosome and the still substantial growth of publicly available databases of companies providing genetic genealogical tests leads to more and more privacy issues which cannot be ignored any longer. Surprisingly, in contrast to 'traditional' ancestry companies providing services to work out genealogies, genetic genealogy companies focus solely on the client and usually make not particularly aware of the implications for third parties, namely

close family members and far-related namesakes (Wallace et al. 2015). Moreover, they also do not stress the possibility of unexpected results which have the potential to be unpleasant for the participant himself. This is however quite important for several reasons. First, it has been reported how difficult it can be for families to accept genetic results when a historically misattributed paternity is discovered or when a persistent family story is contradicted by a simple swab sample (Williams 2005). Such discoveries can even have some serious repercussions for members within noble families or privileged patrilineal groups (Bingham 2016). Second, since the extra-pair paternity rate is 1-2% in human populations, it is of course not uncommon that participants of genetic genealogical tests found out that they are not the father of their supposed child or *vice versa*. Such a discovery or even the possibility of a recent misattributed paternity that may explain a discordant result in their genetic genealogical analysis, can lead to traumatic experiences for participants and their families (Engber 2013; Zarembo 2009). Third, although the incidence is low (namely almost 1 in 4000 males), it is always possible to find an absence of commonly used Y-SNPs and Y-STRs which are associated with male infertility (King et al. 2005). Finally, public databases and websites publishing associations of Y-chromosomal data with surnames has the potential to breach genetic privacy (Erlich and Narayanan 2014). Gymrek *et al.* (2013) succeeded even in identifying anonymised personal genomes of the HapMap project by surname inferences. In combination with other so-called non-identifying information, there are also already several reported cases whereby first-degree relatives were identified in adoption or sperm donor cases (Lehmann-Haupt 2010; Motluk 2005). On the other side of the ethical debate, adoptees, foundlings, GI babies and the donor-conceived have now the possibility using Y-DNA and autosomal DNA testing to find answers about their recent ancestry. Due to the growing popularity of genetic genealogy and ancestry tests, bioethicists have already stated that sperm donor anonymity, which is common in many countries, is under threat (Borry et al. 2014; Borry et al. 2013; Harper et al. 2016). To deal with these ethical concerns, DTC companies should take the effort to inform their clients clearly about those potential implications, which is not sufficient enough until today (Wallace et al. 2015). Fortunately, citizen scientists themselves are becoming more and more aware that they should warn genetic genealogy users about the possibility of unpleasant results and focus on issues such as ethics and privacy, and they have formulated some 'genetic genealogy standards' (see http://www.geneticgenealogystandards.com/). Indeed, the usefulness of publication of particular Y-chromosomal profiles together with surname or paternal lineage data always need to be balanced with the potential ethical repercussions, and this is applicable to researchers as well as to citizen scientists (Larmuseau et al. 2016a).

Conclusion

In 2003, Jobling and Tyler-Smith (2003) wrote their influential paper 'The human Y chromosome: an evolutionary marker comes of age' about the preceding new avenues for investigating human evolution by Y-chromosomal research. In the last decade, the male chromosome fulfilled these promises in numerous evolutionary genetic studies. In 2016, the Y chromosome is just a single locus that is being pushed to the background of human population genetic studies, in favour of full genome analysis. Nonetheless, due to the association with surname and patrilineage, the Y chromosome is still the most popular marker in genetic genealogy. Although genomic studies are commonplace and accessible to most research laboratories, this will remain in the future due to the many unique applications Y-chromosomal data generates when it is combined to genealogical data. In upcoming years, the efficacy of these applications will still increase when more and more full Y chromosome sequences will be available, making tMRCA estimations more accurate, and when genealogical data will become more available due to the exponential growth of digital genealogical and historical demographical databases. Nonetheless, due to its popularity and still increasing possibilities, a thorough bioethical investigation is needed to find an adequate balance between research aims and privacy concerns in Y-chromosomal research.

Conflict of interest

The corresponding authors state that there is no conflict of interest.

Acknowledgements

MHDL is a postdoctoral fellow of the FWO-Vlaanderen (Research Foundation-Flanders). This study is part of the KU Leuven BOF C1-project C12/15/013, the KU Leuven BOF-GOA grant 3H130264, and the Fund for Scientific Research – Flanders grant 1503216N. FC is supported in part by Agència de Gestió d'Ajuts Universitaris i de la Recerca (Generalitat de Catalunya) grant 2014 SGR 866. We thank two anonymous referees and Neus Solé-Morata (Institut de Biologia

Evolutiva, Barcelona), Anke Peeters and Michiel Vandenbosch (KU Leuven) for reading a previous version of the manuscript and helping improve it.

References

Al-Khudhair A et al. (2015) Inference of distant genetic relations in humans using "1000 genomes". Genome Biol Evol 7:481-492

Athey WT (2005) Haplogroup prediction from Y-STR values using an allele-frequency approach. Journal of Genetic Genealogy 1:1-7

Athey WT (2006) Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. Journal of Genetic Genealogy 2:34-39

Baeta M, Nunez C, Cardoso S, Palencia-Madrid L, Herrasti L, Etxeberria F, de Pancorbo MM (2015) Digging up the recent Spanish memory: genetic identification of human remains from mass graves of the Spanish Civil War and posterior dictatorship. Forensic science international Genetics 19:272-279

Balaresque P et al. (2009) Genomic complexity of the Y-STR DYS19: inversions, deletions and founder lineages carrying duplications. International Journal of Legal Medicine 123:15-23

Balaresque P, Poulet N, Cussat-Blanc S, Gerard P, Quintana-Murci L, Heyer E, Jobling MA (2015) Y-chromosome descent clusters and male differential reproductive success: young lineage expansions dominate Asian pastoral nomadic populations. European Journal of Human Genetics 23:1413-1422

Ballantyne KN et al. (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. American Journal of Human Genetics 87:341-353

Ballantyne KN et al. (2012) A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Science International-Genetics 6:208-218

Ballantyne KN et al. (2014) Towards complete male individualization with rapidly mutating Y-chromosomal STRs. Human Mutation 35:1021-1032

Balloux F (2010) The worm in the fruit of the mitochondrial DNA tree. Heredity 104:419-420

Bingham J (2016) Accountant wins legal claim to baronetcy in 'cuckoo in the nest' case that could challenge the British class system. Telegraph Media Group, London.

Bloothooft G, Schraagen MP (2015) Learning name variants from inexact high-confidence matches. In: Bloothooft G, Christen P, Mandemakers K, Schraagen M (eds) Population Reconstruction. Springer, pp 61-83.

Boattini A et al. (2015) Traces of medieval migrations in a socially stratified population from Northern Italy. Evidence from uniparental markers and deep-rooted pedigrees. Heredity 114:155-162

Bolund E, Hayward A, Pettay JE, Lummaa V (2015) Effects of the demographic transition on the genetic variances and covariances of human life-history traits. Evolution 69:747-755

769 Borry P, Rusu O, Dondorp W, De Wert G, Knoppers BM, Howard HC (2014) Anonymity 2.0:
770 direct-to-consumer genetic testing and donor conception. Fertility and Sterility
771 101:630-632
772 Borry P, Rusu O, Howard HC (2013) Genetic testing: Anonymity of sperm donors under threat.
773 Nature 496:169
774 Bosch E et al. (1999) Variation in short tandem repeats is deeply structured by genetic
775 background on the human Y chromosome. American Journal of Human Genetics
776 65:1623-1638
777 Bowden GR et al. (2008) Excavating past population structures by surname-based sampling:
778 The genetic legacy of the Vikings in northwest England. Molecular Biology and
779 Evolution 25:301-309
780 Brown K (2002) Tangled roots? Genetics meets genealogy. Science 295:1634-1635
781 Butler JM (2015) The future of forensic DNA analysis. Philosophical Transactions of the Royal
782 Society B-Biological Sciences 370:20140252
783 Byrnes JK, Myers NM, Underhill PA (2014) Genetic genealogy in the Genomics Era. In: Primorac
784 D, Schanfield M (eds) Forensic DNA Application - An interdisciplinary perspective. Boca
785 Raton, CRC Press, pp 483-508.
786 Calderon R, Hernandez CL, Cuesta P, Dugoujon JM (2015) Surnames and Y-chromosomal
787 markers reveal low relationships in Southern Spain. Plos One 10:e0123098
788 Chaix R, Austerlitz F, Khegay T, Jacquesson S, Hammer M, Heyer E, Quintana-Murci L (2004)
789 The genetic or mythical ancestry of descent groups: lessons from the Y chromosome.
790 American Journal of Human Genetics 75:1113-1116
791 Chaix R, Quintana-Murci L, Khegay T, Hammer M, Mobasher Z, Austerlitz F, Heyer E (2007)
792 From social to genetic structures in Central Asia. Current Biology 17:43-48
793 Charlier P et al. (2013) Genetic comparison of the head of Henri IV and the presumptive blood
794 from Louis XVI (both Kings of France). Forensic Science International-Genetics 226:38-
795 40
796 Chikhi L (2009) Update to Chikhi et al.'s "Clinal variation in the nuclear DNA of Europeans"
797 (1998): genetic data and storytelling--from archaeogenetics to astrologenetics? Hum
798 Biol 81:639-643
799 Colantonio SE, Lasker GW, Kaplan BA, Fuster V (2003) Use of surname models in human
800 population biology: A review of recent developments. Human Biology 75:785-807
801 de Rooij M, M'Charek A (2015) Binnen no time in de Middeleeuwen. Sociologie Magazine
802 2016:22-24
803 Dongying W, Sucan M, Ming L, Shangzhi H, Chunyun L (2000) Application of Y-chromosome
804 polymorphisms to studying the kinship of the males with family name Wang from Han
805 nationality. Acta Anthropologica Sinica 19:132-137
806 Engber D (2013) Who's your daddy? The perils of personal genomics, 21 May 2013 edn.
807 Erlich Y, Narayanan A (2014) Routes for breaching and protecting genetic privacy. Nature
808 Reviews Genetics 15:409-421
809 Foster EA et al. (1998) Jefferson fathered slave's last child. Nature 396:27-28
810 Gillis JR (1985) For Better, for Worse - British Marriages, 1600 to the Present. Oxford University
811 Press, New York.
812 Greeff JM, Erasmus JC (2015) Three hunderd years of low non-paternity in a human
813 population. Heredity 115:396-404
814 Greeff JM, Greeff FA, Greeff AS, Rinken L, Welgemoed DJ, Harris Y (2012) Low nonpaternity
815 rate in an old Afrikaner family. Evolution and Human Behavior 33:268-273
816 Gymrek M, A.L. M, Golan D, Halperin E, Erlich Y (2013) Identifying personal genomes by
817 surname inference. Science 339:321-324
818 Haas C et al. (2013) Y-chromosomal analysis identifies the skeletal remains of Swiss national
819 hero Jörg Jenatsch (1596-1639). Forensic Science International-Genetics 7:610-617

820 Hammer M et al. (2009) Extended Y chromosome haplotypes resolve multiple and unique
821 lineages of the Jewish priesthood. Human Genetics 126:707-717
822 Harper JC, Kennett D, Reisel D (2016) The end of donor anonymity: how genetic testing is likely
823 to drive anonymous gamete donation out of business. Humum Reproduction 31:1135-
824 1140
825 Harris DJ (2016) Does contraceptive use lead to increased affairs? A reply to Larmuseau et al.
826 2016. Trends in Ecology and Evolution 31:663
827 Hayward AD, Lummaa V, Bazykin GA (2015) Fitness consequences of advanced ancestral age
828 over three generations in humans. Plos One 10
829 Helgason A et al. (2015) The Y-chromosome point mutation rate in humans. Nature Genetics
830 47:453-457
831 Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefansson K (2003) A populationwide
832 coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a
833 faster evolutionary rate of mtDNA lineages than Y chromosomes. American Journal of
834 Human Genetics 72:1370-1388
835 Huff CD et al. (2011) Maximum-likelihood estimation of recent shared ancestry (ERSA).
836 Genome Res 21:768-774
837 Hughes JF, Page DC (2015) The biology and evolution of mammalian Y chromosomes. Annual
838 Review of Genetics 49:22.21-22.21
839 Hughes JF, Page DC (2016) The history of the Y chromosome in man. Nat Genet 48:588-589
840 Jehaes E et al. (1998) Mitochondrial DNA analysis on remains of a putative son of Louis XVI,
841 King of France and Marie-Antoinette. European Journal of Human Genetics 6:383-395
842 Jehaes E, Pfeiffer H, Toprak K, Decorte R, Brinkmann B, Cassiman JJ (2001) Mitochondrial DNA
843 analysis of the putative heart of Louis XVII, son of Louis XVI and Marie-Antoinette.
844 European Journal of Human Genetics 9:185-190
845 Jobling MA (2001) In the name of the father: surnames and genetics. Trends in genetics : TIG
846 17:353-357
847 Jobling MA (2008) Copy number variation on the human Y chromosome. Cytogenetic and
848 Genome Research 123:1-4
849 Jobling MA (2012) The impact of recent events on human genetic diversity. Philosophical
850 Transactions of the Royal Society B-Biological Sciences 367:793-799
851 Jobling MA, Hollox E, Hurles ME, Kivisild T, Tyler-Smith C (2013) Human Evolutionary Genetics.
852 Garland Science Publishing, London/New York.
853 Jobling MA, Rasteiro R, Wetton JH (2016) In the blood: the myth and reality of genetic markers
854 of identity. Ethnic Racial Stud 39:142-161
855 Jobling MA, Tyler-Smith C (2003) The human Y chromosome: An evolutionary marker comes of
856 age. Nature Reviews Genetics 4:598-612
857 Johansson MM, Van Geystelen A, Larmuseau MHD, Djurovic S, Andreassen OA, Agartz I, Jazin E
858 (2015) Microarray analysis of copy number variants on the human Y chromosome
859 reveals novel and frequent duplications overrepresented in specific haplogroups. Plos
860 One 10:e0137223
861 Kayser M, Ballantyne KN (2014) Y chromosome in forensic science. In: Primorac D, Shanfield M
862 (eds) Forensic DNA Applications: An Interdisciplinary Perspective. CRC Press Taylor &
863 Francis Group.
864 King TE, Ballereau SJ, Schürer KE, Jobling MA (2006) Genetic signatures of coancestry within
865 surnames. Current Biology 16:384-388
866 King TE, Bosch E, Adams SM, Parkin E, Rosser ZH, Jobling MA (2005) Inadvertent diagnosis of
867 male infertility through genealogical DNA testing. Journal of Medical Genetics 42:366-
868 368
869 King TE et al. (2014) Identification of the remains of King Richard III. Nature communications
870 5:5631

King TE, Jobling MA (2009a) Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. Molecular Biology and Evolution 26:1093-1102
King TE, Jobling MA (2009b) What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. Trends in Genetics 25:351-360
Kottak CP (2002) Cultural Anthropology. Ninth edition edn. Mc Graw Hill, New York.
Lalueza-Fox C, Gigli E, Bini C, Calafell F, Luiselli D, Pelotti S, Pettener D (2011) Genetic analysis of the presumptive blood from Louis XVI, king of France. Forensic Science International-Genetics 5:459-463
Lareu MV, Garcia-Magarinos M, Phillips C, Quintela I, Carracedo A, Salas A (2012) Analysis of a claimed distant relationship in a deficient pedigree using high density SNP data. Forensic science international Genetics 6:350-353
Larmuseau MHD, Bekaert B, Baumers M, Wenseleers T, Deforce D, Borry P, Decorte R (2016a) Biohistorical materials and contemporary privacy concerns - The forensic case of King Albert I. FSI: Genetics 24:202-210
Larmuseau MHD et al. (2015) High Y chromosomal diversity and low relatedness between paternal lineages on a communal scale in the Western European Low Countries during the surname establishment. Heredity 115:3-12
Larmuseau MHD et al. (2014a) Genetic genealogy reveals true Y halogroup of House of Bourbon contradicting recent identification of the presumed remains of two French Kings. European Journal of Human Genetics 22:681-687
Larmuseau MHD, Matthijs K, Wenseleers T (2016b) Cuckolded fathers rare in human populations. Trends in Ecology and Evolution 31:327-329
Larmuseau MHD, Matthijs K, Wenseleers T (2016c) Long-term trends in human extra-pair paternity: increased infidelity or adaptive strategy? A reply to Harris (2016). Trends in Ecology and Evolution 31:663-665
Larmuseau MHD, Ottoni C, Raeymaekers JAM, Vanderheyden N, Larmuseau HFM, Decorte R (2012a) Temporal differentiation across a West-European Y-chromosomal cline - genealogy as a tool in human population genetics. European Journal of Human Genetics 20:434-440
Larmuseau MHD, Van Geystelen A, van Oven M, Decorte R (2013a) Genetic genealogy comes of age - Perspectives on the use of deep-rooted pedigrees in human population genetics. American Journal of Physical Anthropology 150:505-511
Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R (2014b) Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance. Annals of Human Genetics 78:92-103
Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, Kayser M, Decorte R (2014c) Increasing phylogenetic resolution still informative for Y chromosomal studies on West-European populations. Forensic Science International-Genetics 9:179-185
Larmuseau MHD, Vanoverbeke J, Gielis G, Vanderheyden N, Larmuseau HFM, Decorte R (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 et al. (2013b) Low historical rates of cuckoldry in a Western European human population traced by Y-chromosome and genealogical data. Proceedings of the Royal Society B 280:20132400
Laslett P, Oosterveen K, Smith RM (1980) Bastardy and its comparative history: studies in the history of illegitimacy and marital nonconformism in Britain, France, Germany, Sweden, North America, Jamaica, and Japan. Harvard University Press, Cambridge (USA).
Lauwerier R (2012) De oude bewoners van Oldenzaal. Tijdschrift van de Rijksdienst voor het Cultureel Erfgoed 3:10-11
Lehmann-Haupt R (2010) Are sperm donors really anonymous anymore? Slate.

923 Li H et al. (2014) Relationship estimation from whole-genome sequence data. PLoS Genet
924 10:e1004144
925 Marks SJ, Levy H, Martinez-Cadenas C, Montinaro F, Capelli C (2012) Migration distance rather
926 than migration rate explains genetic diversity in human patrilocal groups. Molecular
927 Ecology 21:4958-4969
928 Marlowe FW (2000) Paternal investment and the human mating system. Behavioural Processes
929 51:45-61
930 Martinez-Cadenas C et al. (2016) The relationship between surname frequency and Y
931 chromosome variation in Spain. European Journal of Human Genetics 24:120-128
932 Martinez-Cruz B et al. (2012) Y-chromosome analysis in individuals bearing the Basarab name
933 of the first dynasty of Wallachian kings. Plos One 7:e41803
934 Martinez-Gonzalez LJ et al. (2012) Surname and Y chromosome in Southern Europe: a case
935 study with Colom/Colombo. European Journal of Human Genetics 20:211-216
936 McEvoy B, Bradley DG (2006) Y-chromosomes and the extent of patrilineal ancestry in Irish
937 surnames. Human Genetics 119:212-219
938 McEvoy B, Simms K, Bradley DG (2008) Genetic investigation of the patrilineal kinship structure
939 of early medieval Ireland. American Journal of Physical Anthropology 136:415-422
940 Meates S (2006) How a DNA Project has produced discoveries in the Meates One-Name Study
941 not possible with paper records alone. Journal of One-Name Studies 6:6-10
942 Mendez FL et al. (2013) An African American paternal lineage adds an extremely ancient root
943 to the human Y chromosome phylogenetic tree. American Journal of Human Genetics
944 92:637-637
945 Mensah MA et al. (2014) Pseudoautosomal region 1 length polymorphism in the human
946 population. Plos Genetics 10:e1004578
947 Moore LT, McEvoy B, Cape E, Simms K, Bradley DG (2006) A Y-chromosome signature of
948 hegemony in Gaelic Ireland. American Journal of Human Genetics 78:334-338
949 Motluk A (2005) Anonymous sperm donor traced on internet. New Scientist.
950 Myres NM, Ekins JE, Lin AA, Cavalli-Sforza LL, Woodward SR, Underhill PA (2007) Y-
951 chromosome short tandem repeat DYS458.2 non-consensus alleles occur
952 independently in both binary haplogroups J1-M267 and R1b3-M405. Croatian Medical
953 Journal 48:450-459
954 Niederstätter H et al. (2012) Pasture names with Romance and Slavic roots facilitate dissection
955 of Y chromosome variation in an exclusively German-speaking Alpine region. Plos One
956 7:e41885
957 Nielsen R, Beaumont MA (2009) Statistical inferences in phylogeography. Molecular Ecology
958 18:1034-1047
959 Olalde I et al. (2014) Genomic analysis of the blood attributed to Louis XVI (1754-1793), king of
960 France. Scientific reports 4:4666
961 Ossowski A, Kus M, Kupiec T, Bykowska M, Zielinska G, Jasinski ME, March AL (2016) The Polish
962 genetic database of victims of totalitarianisms. Forensic Science International 258:41-
963 49
964 Pajnic IZ, Pogorelc BG, Balazic J (2010) Molecular genetic identification of skeletal remains
965 from the Second World War Konfin I mass grave in Slovenia. International Journal of
966 Legal Medicine 124:307-317
967 Phillips AD (2016) Only a click away - DTC genetics for ancestry, health, love...and more: A view
968 of the business and regulatory landscape'. Applied & Translational Genomics 8:16-22
969 Pickrell JK, Reich D (2014) Toward a new history and geography of human genes informed by
970 ancient DNA. Trends in Genetics 30:377-389
971 Plant J (2005) Modern methods and a controversial surname: Plant. Nomina 28:115-133
972 Plant J (2007) The tardy adoption of the Plantagenet surname. Nomina 30:57-84
973 Pomery C (2010a) The advantages of a dual DNA/documentary approach to reconstruct the
974 family trees of a surname. Journal of Genetic Genealogy 5:85-95

975 Pomery C (2010b) Defining a methodology to reconstruct the family trees of a surname within
976 a DNA/documentary dual approach project. Journal of Genetic Genealogy 6:1-21
977 Poznik GD et al. (2016) Punctuated bursts in human male demography inferred from 1,244
978 worldwide Y-chromosome sequences. Nat Genet 48:593-599
979 Purps J et al. (2014) A global analysis of Y-chromosomal haplotype diversity for 23 STR loci.
980 Forensic Science International-Genetics 12:12-23
981 Quintana-Murci L, Fellous M (2001) The Human Y Chromosome: The Biological Role of a
982 "Functional Wasteland". Journal of biomedicine & biotechnology 1:18-24
983 Rabb TK, Rotberg RI (1971) The Family in History - Interdisciplinary Essays. Harper Torchbooks,
984 New York.
985 Ralph P, Coop G (2013) The geography of recent genetic ancestry across Europe. Plos Biology
986 11:e1001555
987 Redmonds G, King T, Hey D (2011) Surnames, DNA, & Family History. Oxford University Press,
988 Oxford.
989 Rocca RA, Magoon G, Reynolds DF, Krahn T, Tilroe VO, Boots PMO, Grierson AJ (2012)
990 Discovery of Western European R1b1a2 Y chromosome variants in 1000 Genomes
991 project data: An online community approach. Plos One 7:e41634
992 Rodriguez G (2014) How genealogy became almost as popular as porn. IPC Media.
993 Roewer L, Geppert M, Purps J, Willuweit S (2014) Das Y-Chromosom als forensischer und
994 genealogischer Marker. BIOspektrum 20:627-630
995 Rohde DLT, Olson S, Chang JT (2004) Modelling the recent common ancestry of all living
996 humans. Nature 431:562-566
997 Royal CD, Novembre J, Fullerton SM, Goldstein DB, Long JC, Bamshad MJ, Clark AG (2010)
998 Inferring genetic ancestry: Opportunities, challenges, and implications. American
999 Journal of Human Genetics 86:661-673
1000 Sanchez-Faddeev H et al. (2013) Ancestral stories of Ghanaian Bimoba reflect millenia-old
1001 genetic lineages. Plos One 8:e65690
1002 Schiffels S et al. (2016) Iron Age and Anglo-Saxon genomes from East England reveal British
1003 migration history. Nat Commun 7:10408
1004 Schlecht J, Kaplan ME, Barnard K, Karafet T, Hammer MF, Merchant NC (2008) Machine-
1005 learning approaches for classifying haplogroup from Y chromosome STR data. Plos
1006 Computational Biology 4:e1000093
1007 Scully JL (2014) Naming the dead: DNA-based identification of historical remains as an act of
1008 care. New Genetics and Society 33:313-332
1009 Sear R (2016) Beyond the nuclear family: an evolutionary perspective on parenting. Curr Opin
1010 Psychol 7:98-103
1011 Shriver MD, Kittles RA (2004) Genetic ancestry and the search for personalized genetic
1012 histories. Nature Reviews Genetics 5:611-U613
1013 Sims LM, Garvey D, Ballantyne J (2009) Improved resolution haplogroup G phylogeny in the Y-
1014 chromosome, revealed by a set of newly characterized SNPs. Plos One 4:e5792
1015 Skorecki K et al. (1997) Y chromosomes of Jewish priests. Nature 385:32-32
1016 Solé-Morata N, Bertranpetit J, Comas D, Calafell F (2014) Recent radiation of R-M269 and high
1017 Y-STR haplotype resemblance confirmed. Annals of Human Genetics 78:253-254
1018 Solé-Morata N, Bertranpetit J, Comas D, Calafell F (2015) Y-chromosome diversity in Catalan
1019 surname samples: insights into surname origin and frequency. European Journal of
1020 Human Genetics 23:1549-1557
1021 Speed D, Balding DJ (2015) Relatedness in the post-genomic era: is it still useful? Nature
1022 reviews Genetics 16:33-44
1023 Strassmann BI, Kurapati NT (2016) What explains patrilineal cooperation. Current
1024 Anthropology 57:S118–S130

Strassmann BI, Kurapati NT, Hug BF, Burke EE, Gillespie BW, Karafet TM, Hammer MF (2012) Religion as a means to assure paternity. Proceedings of the National Academy of Sciences of the United States of America 109:9781-9785
Swift C (2013) Interlaced scholarship: genealogies and genetics in twenty-first-century Ireland. In: Duffy S (ed) Princes, Prelates and Poets in Medieval Ireland. Four Courts Press, Dublin, Ireland, pp 19-31.
Sykes B, Irven C (2000) Surnames and the Y chromosome. American Journal of Human Genetics 66:1417-1419
Thomas MG (2013) To claim someone has 'Viking ancestors' is no better than astrology. Guardian News and Media, London.
Tofanelli S, Taglioli L, Bertoncini S, Francalacci P, Klyosov A, Pagani L (2014) Mitochondrial and Y chromosome haplotype motifs as diagnostic markers of Jewish ancestry: a reconsideration. Frontiers in Genetics 5:384
Trombetta B, Cruciani F, Underhill PA, Sellitto D, Scozzari R (2010) Footprints of X-to-Y gene conversion in recent human evolution. Molecular Biology and Evolution 27:714-725
Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annual Review of Genetics 41:539-564
Van Geystelen A, Wenseleers T, Decorte R, Caspers M, Larmuseau MHD (2014) *In silico* detection of phylogenetic informative Y-chromosomal SNPs from WGS data. Electrophoresis 35:3102-3110
van Oven M, Kayser M (2008) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation 30:E386-E394
van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MHD (2014) Seeing the wood for the trees: a minimal reference phylogeny for the human Y chromosome. Human Mutation 35:187-191
Wagner JK, Cooper JD, Sterling R, Royal CD (2012) Tilting at windmills no longer: a data-driven discussion of DTC DNA ancestry tests. Genet Med 14:586-593
Wallace SE, Gourna EG, Nikolova V, Sheehan NA (2015) Family tree and ancestry inference: is there a need for a 'generational' consent? BMC Medical Ethics 16:87
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
Wang CC et al. (2012) Present Y chromosomes reveal the ancestry of Emperor CAO Cao of 1800 years ago. Journal of Human Genetics 57:216-218
Wang CC et al. (2013) Ancient DNA of Emperor CAO Cao's granduncle matches those of his present descendants: a commentary on present Y chromosomes reveal the ancestry of Emperor CAO Cao of 1800 years ago. Journal of Human Genetics 58:238-239
Wilkins JF (2006) Unraveling male and female histories from human genetic data. Current Opinion in Genetics & Development 16:611-617
Willems T, Gymrek M, Poznik GD, Tyler-Smith C, Erlich Y (2016) Population-scale sequencing data enable precise estimates of Y-STR mutation rates. Am J Hum Genet 98:919-933
Williams SR (2005) Genetic genealogy: The Woodson family's experience. Culture Medicine and Psychiatry 29:225-252
Xue Y et al. (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am J Hum Genet 77:1112-1116
Zarembo A (2009) DNA can reveal ancestors' lies and secrets, 18 January 2009 edn., Los Angeles.
Zerjal T et al. (2003) The genetic legacy of the Mongols. American Journal of Human Genetics 72:717-721
Zerubavel E (2012) Ancestors and Relatives - Genealogy, Identity, and Community. Oxford University Press, New York.

1076