

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

3 *Francesc Calafell<sup>1</sup> & Maarten H.D. Larmuseau<sup>2,3</sup>*

4  
5 *<sup>1</sup>Institute of Evolutionary Biology (CSIC-UPF), Departament de Ciències Experimentals i de la*  
6 *Salut, Universitat Pompeu Fabra, Barcelona, Catalonia, Spain*

7 *<sup>2</sup>KU Leuven, Forensic Biomedical Sciences, Department of Imaging & Pathology, Leuven,*  
8 *Belgium*

9 *<sup>3</sup>KU Leuven, Laboratory of Socioecology and Social Evolution, Department of Biology, Leuven,*  
10 *Belgium*

11  
12 **Corresponding authors:**

13 - Prof. dr. Francesc Calafell, Institute of Evolutionary Biology (CSIC-UPF), CEXS-UPF-PRBB,  
14 Doctor Aiguader 88, 08005 Barcelona, Catalonia, Spain. Phone: +34-93.316.08.42; fax: +34-  
15 93.316.09.01; e-mail: francesc.calafell@upf.edu

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

19  
20 **Running title:** Y chromosome popular in genetic genealogy

21  
22 **Keywords:** Genetic genealogy, Y chromosome, Interdisciplinary research, Extra-pair paternity,  
23 Surnames, Patrilineage

24  
25  
26 **Abstract**

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

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

46  
47

## 48 **Introduction**

49

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

64

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

70 can be used to determine the number of meioses that separates two individuals (Lareu et al.  
71 2012). Beyond close family (>6th degree relationships), however, both methods suffer from  
72 low sensitivity and precision even when using the whole genome, due to the vagaries of  
73 recombination (Byrnes et al. 2014; Huff et al. 2011; Li et al. 2014; Speed and Balding 2015).  
74 Recently developed methods based on rare shared variants in whole genome data enlarge the  
75 possibilities to allow detecting relationships of a higher degree (Al-Khudhair et al. 2015;  
76 Schiffels et al. 2016), however, their sensitivity is still poorer than that afforded by haploid  
77 markers. Unilinearly transmitted genome segments, namely mitochondrial DNA (mtDNA) and  
78 the non-recombining portion of the Y chromosome (NRY) are strictly passed down through  
79 respectively the maternal and paternal lines, and could in theory be used to trace them. Note  
80 that, given the lack of recombination in these genome regions, they have been inherited  
81 through the genealogical tree that unites either all human maternal lineages or all paternal  
82 lineages (Underhill and Kivisild 2007).

83

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

100

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

105 by sequencing 1,244 human Y chromosomes from the 1000 Genomes Project (Poznik et al.  
106 2016). Each one of these Y-SNPs first originated at some branch of the genealogical tree that  
107 unites all human males. Some Y-SNPs may have deep origins and are shared by most men,  
108 others may be more recent and are spread in some geographical regions, whereas others may  
109 be so recent that they are confined to a single family or even a single individual. Nevertheless,  
110 their sheer number has allowed to reconstruct the main branches of the paternal phylogenetic  
111 tree which are referred to as *haplogroups*. Population studies have mapped the geographical  
112 deployment of this tree, often referred to as the phylogeographical structure of the Y  
113 chromosome (Jobling and Tyler-Smith 2003). In genetic genealogy, Y-SNPs can thereby be used  
114 to establish the evolutionary lineage and the broad geographic origin of a particular Y  
115 chromosome. Y-STRs are less numerous than Y-SNPs: the most comprehensive catalogue lists  
116 4500 of them (Willems et al. 2016). However, they have faster mutation rates than Y-SNPs: the  
117 average mutation rate for all polymorphic Y-STRs is  $3.83 \times 10^{-4}$  mutations per generation  
118 (Willems et al. 2016), although the most commonly employed and commercially available  
119 subsets have faster mutation rates:  $2.80 \times 10^{-3}$  (www.yhrd.org, accessed June 16th, 2016) for  
120 the 16 most used Y-STRs, and 13 rapid-mutating Y-STRs reach mutation rates  $>10^{-2}$  (Ballantyne  
121 et al. 2012). Combination of Y-STR alleles are referred to as *haplotypes*. Since each Y-STR  
122 mutation happens in some Y chromosome with a particular Y-SNP haplogroup, Y-STR allele  
123 variation is deeply partitioned by haplogroup (Bosch et al. 1999) after which Y-STR haplotypes  
124 can be used with varying precision to predict Y-chromosomal haplogroups in the absence of Y-  
125 SNP genotypes (Athey 2005; Athey 2006; Schlecht et al. 2008). The occurrence of duplicated Y-  
126 STR alleles and intermediate Y-STR allele variants is especially useful in haplogroup predicting  
127 as they are often linked to a specific haplogroup (Balaesque et al. 2009; Myres et al. 2007).

128

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

140 while with a pedigree mutation rate of  $3.07 \times 10^{-8}$  per basepair and generation (Helgason et al.  
141 2015), the probability of having more than 10 differences in a pair of fourth cousins is 0.27%  
142 (estimated under an infinite sites model with a Poisson distribution). Closer relationships  
143 would obviously bear even more similar Y chromosomes. As for Y-STRs, one of the most  
144 popular sets, i.e., the *PowerPlex*<sup>®</sup> Y23 System (Promega, Madison WI, USA), yields a closer call:  
145 0.006% of the pairs among 705 unrelated Spanish males (Purps et al. 2014) carry identical 23-  
146 Y-STR haplotypes, while we estimate that 39.83% of fourth cousin pairs would have identical  
147 STR haplotypes when using locus-specific mutation rates ([www.yhrd.org](http://www.yhrd.org), accessed June 16th,  
148 2016) and a strictly stepwise mutation model (Figure 1). A more definite separation could be  
149 obtained by adding rapidly-mutating Y-STRs (Ballantyne et al. 2014) or Y-SNPs within the R1b  
150 haplogroup (Larmuseau et al. 2014b; Solé-Morata et al. 2014). In general, the combination of  
151 appropriately selected Y-SNPs and large numbers of Y-STRs makes it possible to fine-tune  
152 genetic genealogic testing (Walsh 2001).

153

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

163

164

### 165 ***Genealogy, identification, and the Y chromosome***

166

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

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

186

187 After this success, we discuss two other cases that provide cautionary tales, rather than  
188 positive identifications using Y-chromosomal data. In 2012, a skeleton was excavated at the  
189 presumed site of the Grey Friars friary in Leicester, the last-known resting place of King Richard  
190 III. Archaeological, osteological and radiocarbon dating data were consistent with being his  
191 remains; mtDNA comparison with two living maternal relatives (although separated by 18 and  
192 20 meioses) gave a match. Taking all this in consideration, this was all positive evidence for the  
193 identification of King Richard III (King et al. 2014). But when a set of five paternal relatives  
194 were compared to each other, four shared a Y-STR haplotype, while the fifth did not. The  
195 shared Y haplogroup did not match with that of the King's remnants. Given the overwhelming  
196 evidence for the remains being those of Richard III, King *et al.* (2014) uncovered two historical  
197 cases of false paternity, one in the four generations in the branch leading to the living relative  
198 carrying a different Y-STR haplotype and one in the 18 generations intervening between the  
199 common ancestor of the five living relatives and Richard III. Although male-line relatives are  
200 easier to find from genealogical records and peerage registries, false paternity is a much more  
201 common occurrence than false maternity, and has to be taken into account in identification  
202 studies based on Y-chromosomal data.

203

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

210 only one difference in repeat among the two relics (Charlier et al. 2013; Lalueza-Fox et al.  
211 2011). However, an analysis of three living male relations, belonging to two different lines that  
212 had common Bourbon ancestors which were direct descendants of Henry IV and ancestors of  
213 Louis XVI, showed that all three shared a haplogroup and had similar 38 Y-STR haplotypes,  
214 which were different from those in the relics (Larmuseau et al. 2014a). Moreover, Henry IV  
215 was maternally related to Louis XVII, and a mtDNA sequence recovered from his preserved  
216 heart did not match that from the mummified Henry IV's head (Jehaes et al. 1998; Jehaes et al.  
217 2001). Later, the authors of the original paper on the Louis XVI's relic went on to sequence the  
218 entire genome of the blood on the handkerchief (Olalde et al. 2014). The results placed the  
219 ancestry of the relic in northern Italy (coincidentally, where the relic surfaced), while Louis  
220 XVI's ancestry was mostly central European. Moreover, some phenotypical traits that could be  
221 predicted from the genomic sequence (such as height, obesity and eye colour) did not match  
222 Louis XVI's traits. The authors argued that either the relic was forged or it was heavily  
223 contaminated with other human sources, and in any case, the partial match with Henry IV's  
224 relic was sheer coincidence.

225

226

### 227 ***Confirming genealogical relatedness***

228

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

242

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

245 discrepancy between the social and the biological father) by comparing large numbers of  
246 distantly related men in populations where oral or documental genealogies can be obtained.  
247 Note that this method would not detect non-paternities produced by patrilineal relatives of  
248 the social father. Strassman *et al.* (2012) found that the rate of non-paternity events was 1.8%  
249 per generation in Dogon (Mali) across all religions. It should be taken into account that this  
250 study was based on oral genealogies, which could lead to some errors and an overestimation  
251 of non-paternity rates. Studies based on written genealogies have produced similarly low  
252 rates: 0.9% per generation in Flanders, Belgium across the last 500 years (Larmuseau *et al.*  
253 2013b), which was exactly the same estimated rate for the last 300 years in the Afrikaner  
254 population in South Africa (Greeff and Erasmus 2015) but slightly lower than that in a North  
255 Italian population (1.2% per generation over the last 400 years) (Boattini *et al.* 2015).

256

257

### 258 ***In the name of the father***

259

260 Surnames and the Y chromosome are inherited together through the paternal line in most  
261 cultures with exceptions due to false paternity, adoption and inheritance of the maternal  
262 surname. Actually, surnames can be modelled as alleles of a single locus in the Y chromosome  
263 and have been used as such to investigate inbreeding and gene flow between populations  
264 through measurement of isonymy (that is, sharing surnames) (Colantonio *et al.* 2003).  
265 Currently, the reverse approach is more common: the genetic diversity can be used to  
266 investigate surnames, whether piecemeal in particular cases, or the overall surname system of  
267 a population. In the first scenario, the relatedness of the individuals carrying a surname can be  
268 easily determined by Y-SNPs and Y-STRs, and thus the common origin of that surname can be  
269 established. Some companies provide the public with the genetic tools necessary to  
270 investigate their own surnames, and as we will discuss below, single-surname investigation is  
271 mostly the province of citizen scientists. However, a few such studies in surnames of some  
272 interest have been undertaken in academia. Sykes and Irven (2000) demonstrated for the first  
273 time that indeed different groups of men carrying the Sykes surname had different Y  
274 chromosome genetic profiles, and thus were descendants of different founders of the  
275 surname. In the same year an identical study was carried out in Chinese men bearing the Wang  
276 surname but no differences in Y-chromosomal variants were found with a random Chinese  
277 sample, which was explained by the complex origins of the surname (Dongying *et al.* 2000).  
278 Other surnames have been investigated by their historical relevance. Martínez-González *et al.*  
279 (2012) showed that the Y chromosomes of bearers of the Colombo and Colom surnames

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

292

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

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

322

323

### 324 ***Clans, lineages, castes and other patrilineal institutions***

325

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

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

358

359

### 360 ***Patrilineal relatedness without a priori expectation***

361

362 Even if there is no expectation of patrilineal relatedness, Y-chromosomal analysis can still  
363 detect males sharing a most recent common ancestor who lived in genealogical or historical  
364 times. One intriguing phenomenon is the presence of large descent groups detectable on  
365 population level with a common ancestor who lived in a time period which is still relevant for  
366 genealogists and historians. These cases are assumed to be the result of strong ‘social  
367 selection’ in the past, as genetic drift cannot be the (only) driver for the wide distribution of  
368 that particular Y-chromosomal cluster and as natural selection on the few Y-chromosomal  
369 genes is assumed to be almost non-existent (Balaesque et al. 2015; Zerjal et al. 2003). Thus  
370 strong social selection with a high reproductive success of a certain male and his descendants  
371 is assumed to be the reason for these bursts in the genetic landscape. Several of those large  
372 descent groups with origins in the historical period are already described in Asian populations  
373 (Balaesque et al. 2015), including the ones attributed to descendants of the legendary  
374 Mongolian leader Genghis Khan (c. 1162-1227) (Zerjal et al. 2003) and of Giocangga (died in  
375 1582) linked to the Qing dynasty (1644-1912) (Xue et al. 2005). The attribution of the large  
376 descent groups to both historical figures are based on the distribution and time estimations of  
377 the origin of high successful Y-STR lineages. Moreover, men whose Y chromosomes match  
378 these defined lineages, certainly descend from two prolific male ancestors which were carriers  
379 of the widespread Y-lineages and which had descendants who had a high reproductive success  
380 as well as wealth and socio-political power. These profiles are appropriate for Genghis Khan  
381 and Giocangga but the identities of these ancestors remain unclear without the analysis of  
382 ancient DNA from the attested remains of candidates or a long-dead descendant as ultimate  
383 proof. Next to the eleven examples of successful Y-lineages in the Asian population detected  
384 by Balaesque *et al.* (2015), another successful lineage was found in Ireland. Almost 20% of the

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

398

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

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

427

428

### 429 **Applications of the genealogical analysis of the Y chromosome**

430

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

436

437

### 438 ***Family history***

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

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

461

462 Second, when a 'paper-trail' seems irrefutable but the DNA evidence is contradictory, the  
463 genetic test can provide information about any form of illegitimacy somewhere in the  
464 genealogy which can be further explored by including extra DNA donors in the analysis (Fig.2a).  
465 Therefore, the Y chromosome provides a test for the comparison of legal and biological  
466 genealogy of a certain person (Greeff et al. 2012). Third, the most popular applications of Y-  
467 chromosomal analysis in family history are without any doubt the surname projects (Jobling  
468 2001; King and Jobling 2009b). Not everyone with the same or a similar surname can be linked  
469 to each other in an established genealogy, and therefore (yet) unlinked bearers of a specific  
470 surname are puzzled if their surname is once independently given to different families or if  
471 they have a common paternal ancestor who cannot be traced back due to a lack of archival  
472 records. Therefore, a surname project can lead to the refinement of family trees and the  
473 inclusion of branches. Nevertheless, to be certain about the historical relatedness between  
474 two lineages with the same (or a similar) surname, participants always need to avoid as much  
475 extra-pair paternity and adoptions in the tested lineages as possible (Fig. 2b).

476

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

487

488

489 ***Demography and Anthropology***

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

509

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

518

519

520 ***Forensic sciences***

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

523 lineages (Kayser and Ballantyne 2014). Especially in sexual assault cases, the Y chromosome  
524 provides essential evidence as it is often the only detectable trace of DNA of the offender (Roewer  
525 et al. 2014). Moreover, because male relatives share an identical Y chromosome profile for several  
526 generations, this chromosome is crucial in familial DNA searching or in the prediction of the  
527 surname within a forensic identification case (King and Jobling 2009b). Next, the combination of Y-  
528 chromosomal data and genealogical information can be applied in different ways in forensic  
529 research. First, the different available and future Y-STR kits can be tested to verify which of them  
530 accurately differentiate individuals between and within paternal lineages. Moreover, they can be  
531 used to improve familial DNA searching and surname prediction (Butler 2015). It is assumed that  
532 there has to be different Y-STR kits for different forensic questions, e.g. the 23 Y-STR kit is better for  
533 familial searching (Purps et al. 2014) in contrast to the rapid mutating (RM) Y-STRs which are better  
534 to differentiate males from one single paternal lineage (Ballantyne et al. 2014). Second, a database  
535 with surnames associated with Y-chromosomal genotypes, can be useful within forensic cases with  
536 no matches in a DNA database. As such, it is still possible to predict the surname of a particular  
537 victim or perpetrator. Nevertheless, as the link between surnames and Y-chromosomal data is weak  
538 for common names and not practical for rare names, it is assumed that a database with  
539 intermediate frequency surnames will be most efficient (King and Jobling 2009a; King and Jobling  
540 2009b). Third, genetic genealogy can provide population genetic data of Y-chromosomal markers  
541 based on genealogically unrelated males. It is important that this data is also generated on different  
542 geographical levels, especially on communal scale in rural regions as the chance of uninformative  
543 haplotype matches in a particular region has to be known before a DNA survey is organised in the  
544 close area of a crime scene (Larmuseau et al. 2015). Fourth, genetic genealogical data provides  
545 accurate estimation of mutation rates of (new) Y-chromosomal markers based on (far-)related  
546 males. These analyses still increase the accuracy to estimate the tMRCA between two individuals  
547 (Ballantyne et al. 2010). Finally, specific forensic projects create a huge genetic genealogical  
548 database as one of the few possibilities to identify bodies in mass graves older than 50 years. By  
549 finding Y-chromosomal matches between skeletons and far-related individuals of potential victims,  
550 identification of long-dead individuals can still be realised like in the Fromelles project related to  
551 Australian and British soldiers in World War I (Scully 2014) and the project to identify victims of  
552 communists and Nazi totalitarian regimes in Poland (Ossowski et al. 2016).  
553  
554

555 ***Population genetics***

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

584

585

586 ***Sex chromosome evolution***

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

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

602

603

#### 604 **Huge popularity and drawbacks**

605

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

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

632

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

652

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

659 close family members and far-related namesakes (Wallace et al. 2015). Moreover, they also do  
660 not stress the possibility of unexpected results which have the potential to be unpleasant for  
661 the participant himself. This is however quite important for several reasons. First, it has been  
662 reported how difficult it can be for families to accept genetic results when a historically  
663 misattributed paternity is discovered or when a persistent family story is contradicted by a  
664 simple swab sample (Williams 2005). Such discoveries can even have some serious  
665 repercussions for members within noble families or privileged patrilineal groups (Bingham  
666 2016). Second, since the extra-pair paternity rate is 1-2% in human populations, it is of course  
667 not uncommon that participants of genetic genealogical tests found out that they are not the  
668 father of their supposed child or *vice versa*. Such a discovery or even the possibility of a recent  
669 misattributed paternity that may explain a discordant result in their genetic genealogical  
670 analysis, can lead to traumatic experiences for participants and their families (Engber 2013;  
671 Zarembo 2009). Third, although the incidence is low (namely almost 1 in 4000 males), it is  
672 always possible to find an absence of commonly used Y-SNPs and Y-STRs which are associated  
673 with male infertility (King et al. 2005). Finally, public databases and websites publishing  
674 associations of Y-chromosomal data with surnames has the potential to breach genetic privacy  
675 (Erlich and Narayanan 2014). Gymrek *et al.* (2013) succeeded even in identifying anonymised  
676 personal genomes of the HapMap project by surname inferences. In combination with other  
677 so-called non-identifying information, there are also already several reported cases whereby  
678 first-degree relatives were identified in adoption or sperm donor cases (Lehmann-Haupt 2010;  
679 Motluk 2005). On the other side of the ethical debate, adoptees, foundlings, GI babies and the  
680 donor-conceived have now the possibility using Y-DNA and autosomal DNA testing to find  
681 answers about their recent ancestry. Due to the growing popularity of genetic genealogy and  
682 ancestry tests, bioethicists have already stated that sperm donor anonymity, which is common  
683 in many countries, is under threat (Borry et al. 2014; Borry et al. 2013; Harper et al. 2016). To  
684 deal with these ethical concerns, DTC companies should take the effort to inform their clients  
685 clearly about those potential implications, which is not sufficient enough until today (Wallace  
686 et al. 2015). Fortunately, citizen scientists themselves are becoming more and more aware  
687 that they should warn genetic genealogy users about the possibility of unpleasant results and  
688 focus on issues such as ethics and privacy, and they have formulated some 'genetic genealogy  
689 standards' (see <http://www.geneticgenealogystandards.com/>). Indeed, the usefulness of  
690 publication of particular Y-chromosomal profiles together with surname or paternal lineage  
691 data always need to be balanced with the potential ethical repercussions, and this is applicable  
692 to researchers as well as to citizen scientists (Larmuseau et al. 2016a).

693

694 **Conclusion**

695

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

712

713

714 **Conflict of interest**

715

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

717

718

719 **Acknowledgements**

720

721 MHDH is a postdoctoral fellow of the FWO-Vlaanderen (Research Foundation-Flanders). This  
722 study is part of the KU Leuven BOF C1-project C12/15/013, the KU Leuven BOF-GOA grant  
723 3H130264, and the Fund for Scientific Research – Flanders grant 1503216N. FC is supported in  
724 part by Agència de Gestió d’Ajuts Universitaris i de Recerca (Generalitat de Catalunya) grant  
725 2014 SGR 866. We thank two anonymous referees and Neus Solé-Morata (Institut de Biologia

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

728

729

730

## 731 **References**

732

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

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

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

739 Baeta M, Nunez C, Cardoso S, Palencia-Madrid L, Herrasti L, Etxeberria F, de Pancorbo MM  
740 (2015) Digging up the recent Spanish memory: genetic identification of human remains  
741 from mass graves of the Spanish Civil War and posterior dictatorship. *Forensic science  
742 international Genetics* 19:272-279

743 Balaesque P et al. (2009) Genomic complexity of the Y-STR *DYS19*: inversions, deletions and  
744 founder lineages carrying duplications. *International Journal of Legal Medicine* 123:15-  
745 23

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

750 Ballantyne KN et al. (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics,  
751 molecular bases, and forensic implications. *American Journal of Human Genetics*  
752 87:341-353

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

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

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

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

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

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

767 Bolund E, Hayward A, Pettay JE, Lummaa V (2015) Effects of the demographic transition on the  
768 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, Khedgy 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, Khedgy 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, Sukan 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 hundred 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, Hurler 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

871 King TE, Jobling MA (2009a) Founders, drift, and infidelity: The relationship between Y  
872 chromosome diversity and patrilineal surnames. *Molecular Biology and Evolution*  
873 26:1093-1102

874 King TE, Jobling MA (2009b) What's in a name? Y chromosomes, surnames and the genetic  
875 genealogy revolution. *Trends in Genetics* 25:351-360

876 Kottak CP (2002) *Cultural Anthropology*. Ninth edition edn. Mc Graw Hill, New York.

877 Lalueza-Fox C, Gigli E, Bini C, Calafell F, Luiselli D, Pelotti S, Pettener D (2011) Genetic analysis  
878 of the presumptive blood from Louis XVI, king of France. *Forensic Science*  
879 *International-Genetics* 5:459-463

880 Lareu MV, Garcia-Magarinos M, Phillips C, Quintela I, Carracedo A, Salas A (2012) Analysis of a  
881 claimed distant relationship in a deficient pedigree using high density SNP data.  
882 *Forensic science international Genetics* 6:350-353

883 Larmuseau MHD, Bekaert B, Baumers M, Wenseleers T, Deforce D, Borry P, Decorte R (2016a)  
884 Biohistorical materials and contemporary privacy concerns - The forensic case of King  
885 Albert I. *FSI: Genetics* 24:202-210

886 Larmuseau MHD et al. (2015) High Y chromosomal diversity and low relatedness between  
887 paternal lineages on a communal scale in the Western European Low Countries during  
888 the surname establishment. *Heredity* 115:3-12

889 Larmuseau MHD et al. (2014a) Genetic genealogy reveals true Y halogroup of House of  
890 Bourbon contradicting recent identification of the presumed remains of two French  
891 Kings. *European Journal of Human Genetics* 22:681-687

892 Larmuseau MHD, Matthijs K, Wenseleers T (2016b) Cuckolded fathers rare in human  
893 populations. *Trends in Ecology and Evolution* 31:327-329

894 Larmuseau MHD, Matthijs K, Wenseleers T (2016c) Long-term trends in human extra-pair  
895 paternity: increased infidelity or adaptive strategy? A reply to Harris (2016). *Trends in*  
896 *Ecology and Evolution* 31:663-665

897 Larmuseau MHD, Ottoni C, Raeymaekers JAM, Vanderheyden N, Larmuseau HFM, Decorte R  
898 (2012a) Temporal differentiation across a West-European Y-chromosomal cline -  
899 genealogy as a tool in human population genetics. *European Journal of Human*  
900 *Genetics* 20:434-440

901 Larmuseau MHD, Van Geystelen A, van Oven M, Decorte R (2013a) Genetic genealogy comes  
902 of age - Perspectives on the use of deep-rooted pedigrees in human population  
903 genetics. *American Journal of Physical Anthropology* 150:505-511

904 Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R  
905 (2014b) Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high  
906 Y-STR haplotype resemblance. *Annals of Human Genetics* 78:92-103

907 Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, Kayser M, Decorte R (2014c)  
908 Increasing phylogenetic resolution still informative for Y chromosomal studies on  
909 West-European populations. *Forensic Science International-Genetics* 9:179-185

910 Larmuseau MHD, Vanoverbeke J, Gielis G, Vanderheyden N, Larmuseau HFM, Decorte R  
911 (2012b) In the name of the migrant father - Analysis of surname origin identifies  
912 historic admixture events undetectable from genealogical records. *Heredity* 109:90-95

913 Larmuseau MHD et al. (2013b) Low historical rates of cuckoldry in a Western European human  
914 population traced by Y-chromosome and genealogical data. *Proceedings of the Royal*  
915 *Society B* 280:20132400

916 Laslett P, Oosterveen K, Smith RM (1980) *Bastardy and its comparative history: studies in the*  
917 *history of illegitimacy and marital nonconformism in Britain, France, Germany,*  
918 *Sweden, North America, Jamaica, and Japan.* Harvard University Press, Cambridge  
919 (USA).

920 Lauwerier R (2012) De oude bewoners van Oldenzaal. *Tijdschrift van de Rijksdienst voor het*  
921 *Cultureel Erfgoed* 3:10-11

922 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 patrilineal 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

1025 Strassmann BI, Kurapati NT, Hug BF, Burke EE, Gillespie BW, Karafet TM, Hammer MF (2012)  
1026 Religion as a means to assure paternity. *Proceedings of the National Academy of*  
1027 *Sciences of the United States of America* 109:9781-9785

1028 Swift C (2013) Interlaced scholarship: genealogies and genetics in twenty-first-century Ireland.  
1029 In: Duffy S (ed) *Princes, Prelates and Poets in Medieval Ireland*. Four Courts Press,  
1030 Dublin, Ireland, pp 19-31.

1031 Sykes B, Irven C (2000) Surnames and the Y chromosome. *American Journal of Human Genetics*  
1032 66:1417-1419

1033 Thomas MG (2013) To claim someone has 'Viking ancestors' is no better than astrology.  
1034 *Guardian News and Media*, London.

1035 Tofanelli S, Taglioli L, Bertocchini S, Francalacci P, Klyosov A, Pagani L (2014) Mitochondrial and  
1036 Y chromosome haplotype motifs as diagnostic markers of Jewish ancestry: a  
1037 reconsideration. *Frontiers in Genetics* 5:384

1038 Trombetta B, Cruciani F, Underhill PA, Sellitto D, Scozzari R (2010) Footprints of X-to-Y gene  
1039 conversion in recent human evolution. *Molecular Biology and Evolution* 27:714-725

1040 Underhill PA, Kivisild T (2007) Use of Y chromosome and mitochondrial DNA population  
1041 structure in tracing human migrations. *Annual Review of Genetics* 41:539-564

1042 Van Geystelen A, Wenseleers T, Decorte R, Caspers M, Larmuseau MHD (2014) *In silico*  
1043 detection of phylogenetic informative Y-chromosomal SNPs from WGS data.  
1044 *Electrophoresis* 35:3102-3110

1045 van Oven M, Kayser M (2008) Updated comprehensive phylogenetic tree of global human  
1046 mitochondrial DNA variation. *Human Mutation* 30:E386-E394

1047 van Oven M, Van Geystelen A, Kayser M, Decorte R, Larmuseau MHD (2014) Seeing the wood  
1048 for the trees: a minimal reference phylogeny for the human Y chromosome. *Human*  
1049 *Mutation* 35:187-191

1050 Wagner JK, Cooper JD, Sterling R, Royal CD (2012) Tilting at windmills no longer: a data-driven  
1051 discussion of DTC DNA ancestry tests. *Genet Med* 14:586-593

1052 Wallace SE, Gournas EG, Nikolova V, Sheehan NA (2015) Family tree and ancestry inference: is  
1053 there a need for a 'generational' consent? *BMC Medical Ethics* 16:87

1054 Walsh B (2001) Estimating the time to the most recent common ancestor for the Y  
1055 chromosome or mitochondrial DNA for a pair of individuals. *Genetics* 158:897-912

1056 Wang CC et al. (2012) Present Y chromosomes reveal the ancestry of Emperor CAO Cao of  
1057 1800 years ago. *Journal of Human Genetics* 57:216-218

1058 Wang CC et al. (2013) Ancient DNA of Emperor CAO Cao's granduncle matches those of his  
1059 present descendants: a commentary on present Y chromosomes reveal the ancestry of  
1060 Emperor CAO Cao of 1800 years ago. *Journal of Human Genetics* 58:238-239

1061 Wilkins JF (2006) Unraveling male and female histories from human genetic data. *Current*  
1062 *Opinion in Genetics & Development* 16:611-617

1063 Willems T, Gymrek M, Poznik GD, Tyler-Smith C, Erlich Y (2016) Population-scale sequencing  
1064 data enable precise estimates of Y-STR mutation rates. *Am J Hum Genet* 98:919-933

1065 Williams SR (2005) Genetic genealogy: The Woodson family's experience. *Culture Medicine*  
1066 *and Psychiatry* 29:225-252

1067 Xue Y et al. (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia.  
1068 *Am J Hum Genet* 77:1112-1116

1069 Zarembo A (2009) DNA can reveal ancestors' lies and secrets, 18 January 2009 edn., Los  
1070 Angeles.

1071 Zerjal T et al. (2003) The genetic legacy of the Mongols. *American Journal of Human Genetics*  
1072 72:717-721

1073 Zerubavel E (2012) *Ancestors and Relatives - Genealogy, Identity, and Community*. Oxford  
1074 University Press, New York.

1075

