Detecting past male-mediated expansions using the Y chromosome

Chiara Batini[1] and Mark A. Jobling[2]

[1] Department of Health Sciences, University of Leicester, UK

[2] Department of Genetics, University of Leicester, UK

Correspondence can be addressed to either author:

Dr Chiara Batini, Department of Health Sciences, University of Leicester, University Road, Leicester LE1 7RH, UK

Tel.: +44 (0)116 229 7251

Email: cb334@le.ac.uk

Prof Mark A. Jobling, Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Tel.: +44 116 252 3427   Fax: +44 116 252 3378

Email: maj4@le.ac.uk

Abstract

Males and females display biological differences that lead to a higher variance of offspring number in males, and this is frequently exacerbated in human societies by mating practices, and possibly by past socio-cultural circumstances. This implies that the genetic record might contain the imprint of past male-mediated expansions, which can be investigated by analysing the male-specific region of the Y chromosome (MSY). Here we review studies that have used MSY data to infer such expansions. Sets of short-tandem repeats (STRs) define haplotypes of very low average frequencies, but in a few cases high-frequency haplotypes are observed, forming the core of descent clusters. Estimates of the ages of such clusters, together with geographical information, have been used to propose powerful historical founders, including Genghis Khan, although without direct supporting evidence. Resequencing of multi-megabase segments of MSY have allowed the construction of detailed phylogenies in which branch lengths are proportional to time, leading to the identification of lineage expansions in the last few millennia as well as the more distant past. Comparisons with maternally-inherited mitochondrial DNA sequence data allow the male-specificity of some of these expansions to be demonstrated. These include expansions in Europe in the last ~5 thousand years that may be associated with a cultural shift during the Bronze Age, as well as expansions elsewhere in the world for which explanations from archaeological evidence are not yet clear.

Keywords: Y chromosome; demographic expansion; sex-biased processes; star-like phylogeny

Introduction

It is all too evident that today men and women do things differently, and this has probably been the case throughout the evolution of our species. The different roles and behaviours of males and females in past societies are of great interest to a wide constituency of historians, archaeologists, sociologists and anthropologists, and, as well as their own traditional disciplines, genetics has something to offer in illuminating these matters. Also, there is another discipline that cares a lot about male-specific behaviour: this is forensic science, and in particular forensic genetics. The great majority of violent crimes are committed by men, and therefore analysis of Y-chromosomal DNA is potentially useful in many cases (Jobling et al. 1997). If past male-specific behaviours have had the effect of reducing Y-chromosome diversity in some populations, then this has important implications for how we interpret Y-chromosome profiles in connection with crimes.

In this review we focus on past male-specific expansions and their influences on current patterns of diversity. We start with some general considerations about the Y chromosome and how expansions might be recognised and then review the literature that is based on short-tandem repeat (STR) variation, before surveying more recent work based on resequencing of Y chromosomes.

The Y chromosome as a tool for investigating the demographic history of males

The male-specific region of the Y chromosome (MSY) is passed from father to son without the reshuffling effects of crossing-over. A consequence of this is that all modern MSYs coalesce back to one ancestral sequence at some point in the past. The carrier of this sequence is sometimes referred to as 'Y-chromosomal Adam' (Gibbons 1997), by analogy with his matrilineal counterpart, mitochondrial Eve (Lewin 1987). Apart from its religio-cultural specificity, this name is unfortunate because the Biblical account suggests that the male population has expanded to its current census size of 3.75

billion from one lonely individual. In fact, this Adam, who is estimated from MSY sequence data (Karmin et al. 2015) to have lived around 254 (95% CI 192–307) thousand years ago (KYA), was one of a population of males, of uncertain size, but probably at least three orders of magnitude larger (Wilder et al. 2004).

Biological differences between the sexes, including reproductive lifespan and degree of investment in offspring, suggest that variance in offspring number is likely to be higher for males than for females. Observations of different current human populations generally confirm this (Brown et al. 2009), with the corollary that the effective population size of females is higher than that of males. The ratio of male to female reproductive variance varies considerably between populations (from <1 to >4), with the major determinant being the nature of mating practices: higher variances are generally observed in populations practicing polygyny (males having multiple mates at one time) or serial monogamy (males and females having several mates sequentially), rather than monogamy (a single lifetime mate for both males and females).

These considerations suggest that male-mediated expansions might have occurred in human history associated with social or economic circumstances in which the reproductive success of a subset of males in a population was particularly high. For such expansions to be recognisable in modern MSY distributions, it might be expected that differential reproductive success was transmitted over several generations, so that the sons of successful males were themselves also successful. Such phenomena would also be expected to have particularly significant effects when population sizes were generally small.

In order to recognise male-specific expansions from genetic data, it is necessary to identify some default expectation or general pattern as a comparator. Ideally, this should involve coalescent or forward simulations of population genetic parameters for differently-inherited components of the genome (Wilson Sayres et al. 2014), and then a comparison of real data to these simulated data (Webster and Wilson Sayres 2016). Under simple neutral expectations MSY diversity is expected to be one quarter of that of autosomal loci, and similar to that of mtDNA (Caballero 1995). Comparisons with mtDNA are expected to show the most divergent patterns since this locus is maternally inherited, but comparisons with X chromosomes (e.g. (Maisano Delser et al. 2017))

(inherited twice as often from mothers as from fathers) or autosomes (equally biparentally inherited) could also be informative (Wilson Sayres et al. 2014). However, in practice proposals of male-specific expansions have often been based on surveys of MSY diversity in a set of samples, and identification of an unusual pattern in a particular population(s) compared to the majority (Zerjal et al. 2003). In all these comparisons, there is an inherent assumption that successful MSY lineages succeed because of social or cultural selection (Chiaroni et al. 2009; Creanza and Feldman 2016), rather than positive natural selection. This seems reasonable: the role of purifying selection on MSY genes has been quite well established by comparisons of primate MSY gene content (Hughes et al. 2012), frequency analysis of potentially damaging genic variants (Hallast et al. 2015; Poznik et al. 2016; Rozen et al. 2009; Scozzari et al. 2014), and simulation approaches (Wilson Sayres et al. 2014), but no evidence has yet been presented for natural selection favouring particular MSY lineages.

Inferences about male lineage expansions from Y-STR analysis

While most early population studies of Y-STRs focused on their forensic informativeness (de Knijff et al. 1997), one demographic modelling study (Pritchard et al. 1999) analysed global samples with eight Y-STRs and noted that a period of recent population growth fitted observed data far better than a constant-sized population model with no growth, pointing to the importance of recent Y-chromosomal expansions.

When a set of 10-20 polymorphic Y-STRs is analysed in populations of males, most haplotypes are unique within a sample of >100 individuals (Alves et al. 2003), a property that makes such markers forensically useful. Consequently, when a particular haplotype is observed at high frequency in population samples this attracts attention and becomes the focus of investigation. In practice, a frequent core haplotype is generally accompanied by a set of closely related haplotypes that may have arisen from it via mutation, constituting a feature referred to as a 'star cluster' (Zerjal et al. 2003), or 'descent cluster' (King and Jobling 2009). Such a cluster is assumed to result from a lineage-specific expansion at some point in the past (Balaresque et al. 2015). Consistent with descent from a recent common ancestor, haplotypes comprising a given cluster generally belong to a particular Y-SNP (single nucleotide polymorphism)-defined haplogroup. Consideration of the total number of mutational steps encompassed by such clusters and the frequencies of constituent haplotypes, together with Y-STR mutation rate estimates and a single-step mutational model, can allow the time to most recent common ancestor (TMRCA) to be estimated. Cluster-based approaches, by their very nature, tend to be focused on lineage-specific expansions, rather than male-mediated expansions that affected an entire past population.

Three descent clusters have been studied in detail and, in each case, a powerful historical man suggested as its source; phylogenetic network analyses enable a graphical representation of the clusters and their relationships with other lineages in the respective populations (Figure 1), and also illustrate the point that the discriminating power of forensic Y-STR analysis in cluster-containing populations is compromised.

(i) *The 'Khan' cluster*:

In a study using 16 Y-STRs to assess MSY diversity in 2123 Asian males (Zerjal et al. 2003), >90% of haplotypes were of very low frequency, but one haplotype and its close mutational neighbours (within the Y-SNP-defined haplogroup C[xC3c], also known as C[xC-M48]) were present at a combined frequency of ~8% in the sample. This 'star cluster' (Figure 1a) was notable not only because of its high frequency, but also because of its wide geographical range, in populations from the Caspian Sea to the Pacific. Three features suggested that expansion of this lineage might have been due to Genghis Khan and his descendants: first, the geographical distribution of the cluster coincided well with the historically recorded range of Khan's empire; second, the population carrying the highest frequency of star cluster haplotypes, which may therefore reflect the geographical origin of the expansion, was Mongolia, consistent with the focus of Khan's Mongol empire; and third, the TMRCA of the cluster was estimated as ~860-1000 years (depending on the method used), with 95% confidence intervals of ~600-1300 years, compatible with an origin in the lifetime of Genghis Khan (c. 1162–1227). The haplogroup within which the star cluster lies has been somewhat refined more recently (Abilev et al. 2012) to hg C3(xC3a-f), using the nomenclature of (Karafet et al. 2008).

(ii) *The 'Manchu' cluster*:

A second study of Asian MSY diversity (Xue et al. 2005) surveyed 1003 males using 15 Y-STRs, identifying a frequent haplotype cluster (Figure 1b) within haplogroup (hg) C-M48 (also known as C3c (Karafet et al. 2008)) constituting ~3% of the sample and present in seven populations (Xibe, Outer Mongolians, Inner Mongolians, Ewenki, Oroqen, Manchu and Hezhe). All of these reside in Mongolia or NE China, with the exception of the western Chinese Xibe who migrated from the north-east of the country ~250 years ago. In this case, increasing the number of STRs from 15 to 61 allowed the boundary of the cluster to be defined more clearly (Figure 1b) and excluded some haplotypes that were previously included at lower Y-STR resolution. The TMRCA of the cluster was estimated as 590±340 years, and it was reasoned that it probably had its origin in the Manchu conquest of China and the establishment of the Qing dynasty. This dynasty was founded by descendants of the historical figure Giocangga, who died in

1582, and its nobility who were highly privileged, with access to many concubines that may have allowed the associated MSY type to expand within a limited number of Asian populations.

(iii) *The 'Uí Néill' cluster*:

In the final example, a sample of 796 Irish males was surveyed using SNPs and a set of 17 Y-STRs (Moore et al. 2006). More than 85% of the MSY haplotypes belonged to hg R1b-M269, but within this haplogroup one Y-STR haplotype stood out as being particularly common in the northwest of Ireland, forming the core of a starlike cluster (Figure 1c) comprising ~21% of the haplotypes in the region, and with a TMRCA of 1010±390 years. Together with historical information supporting the existence of patrilineal, polygynous social structures in medieval Ireland, a link was made between this cluster and the Uí Néill, a powerful Irish medieval royal lineage allegedly descending from Niall of the Nine Hostages, a possibly mythological 5th-century warlord. The significant over-representation of the cluster haplotypes among men bearing surnames thought to descent from the Uí Néill lineage was taken to support the connection.

These claimed links to powerful historical patriarchs have strong cultural resonance, as attested by the offer of Genghis Khan or Niall of the Nine Hostages DNA tests that have been offered by direct-to-consumer testing companies. However, the three studies described above provide no definitive proof that the expansions they identify are linked to their claimed historical origins, but instead a range of circumstantial evidence. Sequences from attested living descendants, or from ancient DNA extracted from attested remains, have not yet been obtained. Ancient DNA data (Lkhagvasuren et al. 2016) from remains in high-status Mongolian graves dated to 1130–1250 CE revealed MSY lineages belonging to hg R1b, rather than hg C: there are a number of explanations for such findings, but taken at face value they do not support the Genghis Khan hypothesis for the origin of the widespread Asian expansion lineage (Zerjal et al. 2003).

Following on from the identification of the two Asian descent clusters described above (Xue et al. 2005; Zerjal et al. 2003), a larger survey was undertaken of 5321 males from 127 Asian populations, based on 8 Y-STRs, to ask if further examples of Asian male expansion lineages could be found (Balaresque et al. 2015). This exercise identified

11 clusters, each within a particular haplogroup, including the two cases identified previously (Xue et al. 2005; Zerjal et al. 2003). Historical founders were not proposed for the additional descent clusters, but an analysis of the ages and population distributions of the expansion lineages suggested that the younger examples were over-represented among Altaic-speaking pastoral nomadic populations, perhaps reflecting a socio-economic shift towards patrilineal descent rules that led to a cultural transmission of reproductive success. This study, and by extension the previous case studies (Moore et al. 2006; Xue et al. 2005; Zerjal et al. 2003), has been challenged by a coalescent-simulation-based approach (Guillot and Cox 2015) showing that frequent Y-STR lineages could appear by chance without the need for transmitted reproductive success. Such lineages were therefore claimed to be not unusual. Although this coalescent study did not take account of the population distributions of simulated lineages, it does underline the need to test observed patterns against neutral expectations as far as is possible.

These various studies, together with others, suggest some general observations about the use of Y-STR haplotypes to infer past male-mediated expansions: (i) Given the high rates of mutation of Y-STRs (typically ~0.1% per STR per generation) (Gusmão et al. 2005), the power of STR-based haplotypes to identify past expansions has a limited time window, and they are most suited to detecting the consequences of events in the last few millennia; (ii) Small numbers of Y-STRs (≤10) are unlikely to define descent clusters clearly against a background of more distantly related haplotypes, so larger numbers (>15) are advisable, as is illustrated by the case of the Manchu cluster (Xue et al. 2005) (Figure 1b); (iii) Criteria for defining haplotype membership of descent clusters vary from study to study (Balaresque et al. 2015; Moore et al. 2006; Xue et al. 2005; Zerjal et al. 2003), and some consistent objective standards would be helpful; (iv) Use of simulations to assess the significance of findings (Guillot and Cox 2015) is a good idea and might also be helpful in establishing cluster definitions; (v) TMRCA estimates are uncertain, due to different methods of estimation, and more importantly, widely differing mutation rates between studies: there is a three-fold difference between the slow 'evolutionary' rate of Y-STR mutation (Zhivotovsky et al. 2004), and the rate established in father-son pairs (Gusmão et al. 2005), though comparisons with MSY resequencing data suggest the latter rate is more accurate for lineages with TMRCAs up to a few

thousand years (Hallast et al. 2015); (vi) Access to aDNA data to calibrate mutation rates or identify putative founders of modern expansion lineages may be limited, given that fragment sizes needed to type Y-STRs are generally >100 bp.

In the discussion above we have focused on studies that identified high-frequency haplotypes and sought to explain their origins. However, there are many other studies that use Y-STRs together with Y-SNPs to analyse population samples, and observe portions of STR networks that are more star-like than others, albeit lacking descent clusters of the kinds represented in Figure 1. Some of these studies have inferred male-mediated expansions from such data. An example is the observation of a star-like network structure within the young (~6 KYA) hg E-M180 among Bantu-speaking farming populations of sub-Saharan Africa (Berniell-Lee et al. 2009). This is probably related to the spread of agriculture and the emergence of iron technologies associated with the so-called 'Bantu expansion', and the fact that mtDNA lineages in the same populations are predominantly much more ancient suggests a male-mediated demographic process underlying the cultural transition. We will not review the extensive Y-STR-based literature here, because resequencing approaches, which we discuss below, are now providing the best insights into these population phenomena.

Inferences based on resequencing of Y chromosomes

Prior to the introduction of Next-Generation Sequencing (NGS) technologies, STR-based analysis was the only approach that could provide evidence of past expansions of MSY lineages. Slowly mutating markers (mostly SNPs) were available (Karafet et al. 2008), but in small numbers and subject to strong ascertainment bias. The field was transformed when an unbiased picture of MSY sequences became available through the introduction of NGS: the rapid decrease in cost and increase in output stimulated a burst of studies (Batini et al. 2015; Francalacci et al. 2013; Hallast et al. 2015; Karmin et al. 2015; Lippold et al. 2014; Poznik et al. 2013; Poznik et al. 2016; Scozzari et al. 2014; Wei et al. 2013; Yan et al. 2014) which could be classed as a dramatic expansion itself, if these were considered as evolutionary units.

As with Y-STRs, detecting an expansion using MSY sequence data relies heavily on reconstructing a phylogeny, investigating its overall topology and suggesting explanations for the relative lengths of internal and external branches. The advantages of using single-nucleotide variants rather than STRs are several: unbiased ascertainment of thousands of SNPs (more than 60,000 have now been reported (Poznik et al. 2016)) allows the reconstruction of unbiased phylogenies with branch lengths expected to be proportional to time; the lower probability of recurrent mutation makes this reconstruction robust, and the slower mutation rate allows the detection of both recent and more ancient events. Thus, NGS-based analysis offers not only the possibility of accurately mapping the structure of MSY clades, with the potential to detect star-like expansion features, but also provides a new temporal depth to the inference of the Y-chromosome evolutionary history.

The higher molecular resolution, together with the possibility of dating ancient events, has provided a much deeper insight into the structure of the Y chromosome evolutionary tree and the demographic dynamics behind it. Leaving aside the implications of these results for deep-rooting lineages and phylogeographic inference, we focus here on evidence of demographic expansions in the distant and recent past. We divide the latest results into early world-wide processes and more recent regional events.

(i) Early expansion after the Out of Africa migration

General agreement has largely been reached on the origin of our species ~200 KYA in sub-Saharan Africa (Jobling et al. 2014), but the number and direction of routes of dispersal outside the continent, towards Asia and Europe, are still somewhat a matter of debate (Nielsen et al. 2017). Deep sequencing of ~10 Mb of MSY, representing its uniquely mappable regions (Poznik et al. 2013), not only confirms the sub-Saharan African origin of MSY diversity, but also supports an early demographic expansion 10-20 KYA after the exit from Africa. In fact, when reconstructing MSY phylogenies, different studies report a sudden ramification of the clades F-R (Figure 2a), which is represented by many short internal branches giving rise to several major clades in a short amount of time (Hallast et al. 2015; Karmin et al. 2015; Poznik et al. 2013; Poznik et al. 2016; Wei et al. 2013). This explosion of lineages is dated at around ~50 KYA by most studies (discrepancies between studies are explained by differences in the mutation rates chosen for calibration). The inference of expansion is based on the high concentration of coalescence events at a certain depth of a tree that contains many haplotypes deriving from many different populations, and in some cases additional haplotypes chosen in a biased manner (Hallast et al. 2015; Poznik et al. 2013; Wei et al. 2013), so it does not necessarily mirror population dynamics. However, when a population approach is applied to macro-regions in a world-wide context (Figure 2b), all groups show an increase of effective population size at ~50 KYA in agreement with the tree (Karmin et al. 2015).

While it seems clear from MSY data that an expansion occurred early after the Out of Africa migration, this does not imply that the process was male-mediated. Nonetheless, comparison with mitochondrial DNA (mtDNA) does suggest sex-biased behaviours during this period, since female effective population size at 50 KYA is about four times larger than that of males, and mtDNA shows signs of expansion at least 10 KY earlier than MSY (Karmin et al. 2015). However, lack of independent sources of evidence for this period make questions on the social structure of early *Homo sapiens* communities currently difficult to address further.

(ii) Regional recent expansions

Recent regional demographic processes can, by contrast, be considered in the

wider context of evidence from other disciplines, and signals of expansion detected from genetic data can be compared with knowledge on population migration and social structure inferred from archaeological and linguistic information.

One example comes from the calibration of the phylogenetic tree obtained from low-coverage MSY sequence data from ~1200 Sardinian men (Francalacci et al. 2013). While mtDNA data have shown that the most frequent Sardinian lineages (including hg H) are common in the rest of Europe (Pardo et al. 2012), the spectrum of MSY lineages is very distinctive. The high frequency (~40%) of hg I2a1a1, rare elsewhere (Rootsi et al. 2004), and its shallow phylogenetic topology as demonstrated by NGS data (Francalacci et al. 2013) have been taken as evidence that this lineage arrived with the initial colonisation of the island ~7.7 KYA, and subsequently expanded locally with other lineages during the Late Neolithic and later periods of cultural innovation.

The mechanisms of spread of cultural innovations over wider geographical areas have been the cause of considerable debate. Classic examples of this are the spread of agriculture in Europe during the Neolithic period (beginning ~10 KYA), and the diffusion of cultural innovations and languages in sub-Saharan Africa during the 'Bantu expansion' (beginning ~5 KYA), mentioned above. In both cases the debated issue is whether these diffusions were mediated by populations migrating to new territories and replacing previous inhabitants (demic diffusion), or by technological innovations spreading without the movement of people (cultural diffusion). This is a topic that could form the subject of a whole review itself (Jobling et al. 2014), and the answer is probably something in between (Pinhasi et al. 2012), but thanks to NGS data the debate has taken unexpected turns.

Initial interpretations of limited Y-STR and Y-SNP data in Europe favoured a Palaeolithic origin for most of the major lineages (Semino et al. 2000), and the star-like pattern of Y-STR variation within the most frequent European MSY clade, hg R1b-M269, was initially interpreted as the result of Neolithic expansion (Balaresque et al. 2010). NGS data have demonstrated that this expansion is even more recent (Batini et al. 2015; Poznik et al. 2016). One study based on the resequencing of 3.7 Mb of MSY (Batini et al. 2015), has shown that three of the most frequent haplogroups (I1, R1a and R1b) show shallow star-like phylogenies and recent TMRCAs (3-7 KYA). The most

frequent, hg R1b-M269, shows a remarkable star-like structure that includes 44 sequences descending from a single node. Thirteen of the 17 individual populations included in this study showed strikingly similar demographic histories characterised by an expansion dated at 2-4 KYA (Batini et al. 2015). Taking into account the much earlier Palaeolithic growth observed in histories inferred from mtDNA variation (Batini et al. 2016; Lippold et al. 2014), these expansions are best explained by a male-specific behaviour which allowed a few MSY lineages to increase greatly in frequency, and diversity, in a short time. This may have been associated with technological advances in the Bronze Age: genome-wide aDNA evidence (Allentoft et al. 2015; Haak et al. 2015) has suggested a massive migration from the Eurasian Steppe during this period, associated with the spread of the Yamnaya culture, and the dispersal of hg R1b-M269 (Cassidy et al. 2015), and the male-bias of steppe migration is supported by modelling of X-chromosomal versus autosomal aDNA data (Goldberg et al. 2017). However, the Western European R1b expansion sub-lineages are not observed in ancient Yamnaya haplotypes obtained to date (Balanovsky et al. 2017), highlighting the incompleteness of our current understanding of expansions during the Bronze Age.

A slightly different picture is presented by African lineages, and specifically by hg E1b-M2, previously suggested to have spread during the expansion of Bantu-speaking populations (Berniell-Lee et al. 2009). No NGS-based study has yet focused on the internal structure of this clade, but the topology resulting from the few individuals analysed in more general phylogenetic reconstructions suggests interesting differences with the topology of hg R1b-M269. Both clades are shallow, but hg E1b-M2 appears notably less star-like, perhaps indicating an expansion from a group with a bigger effective population size (Wei et al. 2013). This more complex structure is confirmed by inspection of phylogenies in later studies (Hallast et al. 2015; Poznik et al. 2013; Trombetta et al. 2015) and shows a deeper coalescence (Hallast et al. 2015; Trombetta et al. 2015). Understanding the implications of these results for the initial movements of Bantu-speaking populations will have to await a dedicated larger-scale study, with a sampling strategy aimed at analysing the history of populations, rather than lineages.

Finally, the first world-wide approach to the detection of regional expansions has been applied recently through an analysis of MSY data in the 1000 Genomes Project

samples (Poznik et al. 2016). Clades showing signals of possible expansion were first identified by examining their topologies, as done in previous studies. Instead of employing Bayesian skyline plots to trace the history of effective population size (Figure 2b), a modelling approach was used to further investigate these potential localised bursts in the male demographic history. This considers two phases of diversification - a rapid expansion followed by moderate growth. Results were presented for eight clades, representing five world-wide macro-regions, and interpreted in the light of historical information (Poznik et al. 2016). The expansion observed within hg Q1a-M3 in the Americas at ~15 KYA was taken as a validation of the method and the temporal calibration used, since it corresponded well with archaeological evidence for the initial colonisation of the New World (Raghavan et al. 2015). More recent expansions were inferred in the other regions: in Europe hg R1b-L11 shows signals of rapid diversification dated to ~5 KYA, in agreement with previous studies (Batini et al. 2015). Inference of an expansion in sub-Saharan Africa was only indirect before this analysis, but here two lineages within hg E1b-M180 were shown to expand at ~5 KYA; it is possible that the expansion pre-dates the Bantu migration, since some of the expansion haplotypes were observed in non-Bantu-speaking populations. Finally, three clades were suggested to be affected by rapid demographic growth in South Asia at 5-6 KYA (hgs H1-M52, L-M11 and R1a-Z93), as well as two hg O sub-haplogroups (hgs O2b-M176 and O3-M122) in East Asia at ~4 and ~8 KYA respectively. It is difficult to relate these changes to events in the archaeological record, but interesting to observe a parallel between different continents in post-Neolithic expansion. This is consistent with the population-based approach illustrated in Figure 2b (Karmin et al. 2015), where all macro-regions analysed show a bottleneck starting ~10 KYA followed by an expansion from ~5-6 KYA, possibly caused by a cultural or technological change that is currently poorly understood.

Perspectives

The generation of phylogenies from NGS data, and availability of comparative data from other parts of the genome, is already revolutionising our understanding of the past roles of sex-biased behaviour, and of male-mediated expansions in particular. These data will continue to accumulate and provided the same regions of the MSY are included

in emerging studies, should be able to be combined into trees containing very large numbers of MSY sequences from a widening range of populations. Some geographical regions are currently understudied, and this includes some populations, from Central Asia and the Middle East, for example, that might be expected to show strong signatures of polygyny and patrilineal descent rules. With larger sample sizes it should be possible to identify past population-based processes affecting multiple lineages, as well as lineage-specific expansions.

There is still some disagreement about the choice of the most appropriate Y-SNP mutation rate for calibrating phylogenies, and this affects interpretations of the TMRCAs of expansions. Rates have been proposed from direct resequencing data in father-son pairs (Helgason et al. 2015), from times of colonisation of particular regions from archaeological data (Francalacci et al. 2013; Poznik et al. 2013), and from MSY sequencing of well-dated ancient remains (Fu et al. 2014). Although the confidence intervals of these estimates overlap, it would help the field if a consensus were to emerge.

In inferring past expansions from MSY sequences, ancient DNA data could be helpful by demonstrating the past geographical ranges, frequencies and diversities of lineages. Current ancient MSY data are limited, but are expected to accumulate, albeit with a bias towards regions in which conditions allow good DNA preservation. Investigation of the Bantu expansion from ancient DNA data, for example, will be challenging.

As noted above, recovering Y-STR alleles from ancient samples is often not possible, but the mass of modern Y-STR data will surely continue to grow, through the efforts of forensic geneticists and amateur genetic genealogists, with their allies the direct-to-consumer DNA testing companies. The number of Y-STRs now available, thanks to MSY resequencing efforts, is several hundred (Willems et al. 2016), and these include so-called rapidly-mutating STRs that promise to provide exquisite resolution to the dissection of descent-clusters (Ballantyne et al. 2010; Ballantyne et al. 2012). Linking highly-resolving Y-STR haplotypes systematically to MSY sequences will in turn provide insights into the relative diversities and mutation rates of these systems, as has been attempted already on a limited scale (Hallast et al. 2015).

Given the enormous quantity of DNA data that promises to emerge over the next decade, much will certainly be learned about male-mediated expansions: the future is expanding.

Acknowledgements

Conflict of Interest Statement

The authors declare no conflict of interest.

Figure legends

Figure 1. Examples of patrilineal descent clusters detected using Y-STR analysis.

Median-joining networks are shown in which circles represent Y-STR haplotypes, with areas proportional to the number of carriers, and lengths of lines between them are proportional to the number of Y-STR mutational steps, as shown in the scale keys in each panel. (a) The 'Khan' cluster (Zerjal et al. 2003) within haplogroup C(xC-M48), based on 15 Y-STRs: the boundary of the identified cluster is shown by the dotted blue line, and populations are indicated by colours as indicated in the key above. Figure redrawn from Zerjal et al. (2003). (b) The 'Manchu' cluster (Xue et al. 2005) within haplogroup C-M48: the upper network is based on 15 Y-STRs, and the lower network on 61 Y-STRs. In each case the boundary of the identified cluster is shown by the dotted blue line, and populations are indicated by colours as indicated in the key above. Figure redrawn from Xue et al. (2005). (c) The 'Uí Néill' cluster (Moore et al. 2006) within haplogroup R1b-M269, based on 17 Y-STRs: all haplotypes are derived from northwest Irish males, and haplotypes included in the cluster are indicated in red. Figure redrawn from Moore et al. (2006).

Figure 2. MSY phylogeny from NGS data, and associated demographic reconstruction.

a) MSY phylogeny based on 456 samples and 35,700 SNPs. Major haplogroups are labelled. The orange box highlights the recent expansions identified in several haplogroups, and the yellow box highlights the more ancient expansion of deep-rooting lineages discussed in the text. b) MSY Bayesian Skyline Plots (of effective population size against time), with different world regions indicated by colours as shown in the key. Figure adapted from Karmin et al. (2015).

References

Abilev S, Malyarchuk B, Derenko M, Wozniak M, Grzybowski T, Zakharov I (2012) The Y-chromosome C3* star-cluster attributed to Genghis Khan's descendants is present at high frequency in the Kerey clan from Kazakhstan. Hum Biol 84: 79-89.

Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlstrom T, Vinner L, Malaspinas AS, Margaryan A, Higham T, Chivall D, Lynnerup N, Harvig L, Baron J, Della Casa P, Dabrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolar J, Kriiska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Palfi G, Pokutta D, Pospieszny L, Price TD, Saag L, Sablin M, Shishlina N, Smrcka V, Soenov VI, Szeverenyi V, Toth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Ponten T, Brunak S, Nielsen R, Kristiansen K, Willerslev E (2015) Population genomics of Bronze Age Eurasia. Nature 522: 167-72.

Alves C, Gusmao L, Barbosa J, Amorim A (2003) Evaluating the informative power of Y-STRs: a comparative study using European and new African haplotype data. Forensic Sci Int 134: 126-33.

Balanovsky O, Chukhryaeva M, Zaporozhchenko V, Urasin V, Zhabagin M, Hovhannisyan A, Agdzhoyan A, Dibirova K, Kuznetsova M, Koshel S, Pocheshkhova E, Alborova I, Skhalyakho R, Utevska O, Genographic Consortium, Mustafin K, Yepiskoposyan L, Tyler-Smith C, Balanovska E (2017) Genetic differentiation between upland and lowland populations shapes the Y‐chromosomal landscape of West Asia. Hum. Genet. in press: DOI 10.1007/s00439-017-1770-2.

Balaresque P, Bowden GR, Adams SM, Leung H-Y, King TE, Rosser ZH, Goodwin J, Moisan J-P, Richard C, Demaine AG, Wilson IJ, Tyler-Smith C, Jobling MA (2010) A predominantly Neolithic origin for European paternal lineages. PLoS Biol. 8: e1000285.

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. Eur J Hum Genet 23: 1413-22.

Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, Decorte R, Poetsch M, von Wurmb-Schwark N, de Knijff P, Labuda D, Vezina H, Knoblauch H, Lessig R, Roewer L, Ploski R, Dobosz T, Henke L, Henke J, Furtado MR, Kayser M (2010) Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. Am J Hum Genet 87: 341-53.

Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M (2012) A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. Forensic Sci Int Genet 6: 208-18.

Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, Ghirotto S, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Dupuy BM, Eriksen HA, King TE, de Munain AL, Lopez-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Tolun A, Winney B, Jobling MA (2015) Large-scale recent expansion of European patrilineages shown by population resequencing. Nat Commun 6: 7152.

Batini C, Hallast P, Zadik D, Vågene A, Boquete L, Jobling MA (2016) Population-level sequencing of European mitochondrial DNAs confirms the male-specificiity of Bronze-Age expansion. in preparation.

Berniell-Lee G, Calafell F, Bosch E, Heyer E, Sica L, Mouguiama-Daouda P, van der Veen L, Hombert JM, Quintana-Murci L, Comas D (2009) Genetic and demographic implications of the Bantu expansion: insights from human paternal lineages. Mol Biol Evol 26: 1581-9.

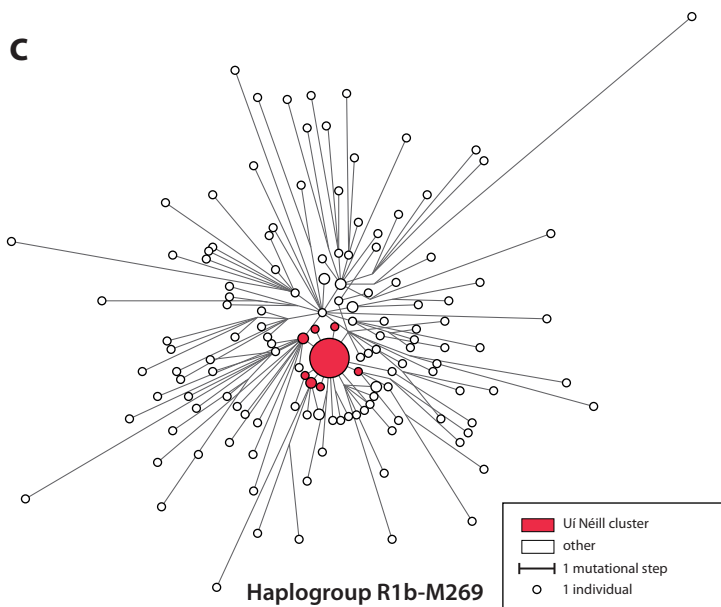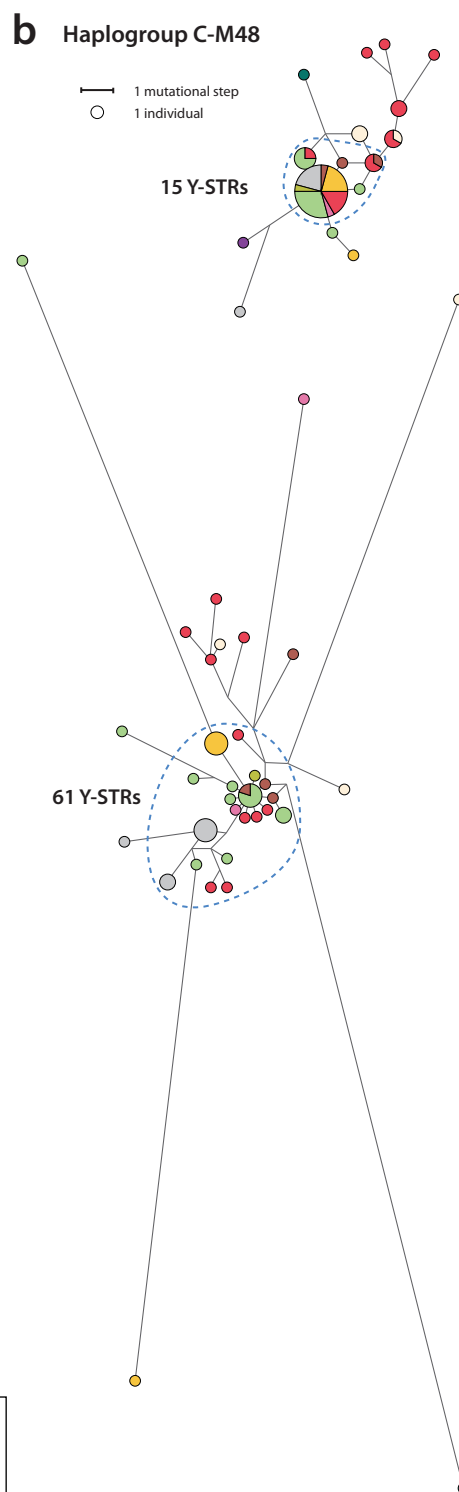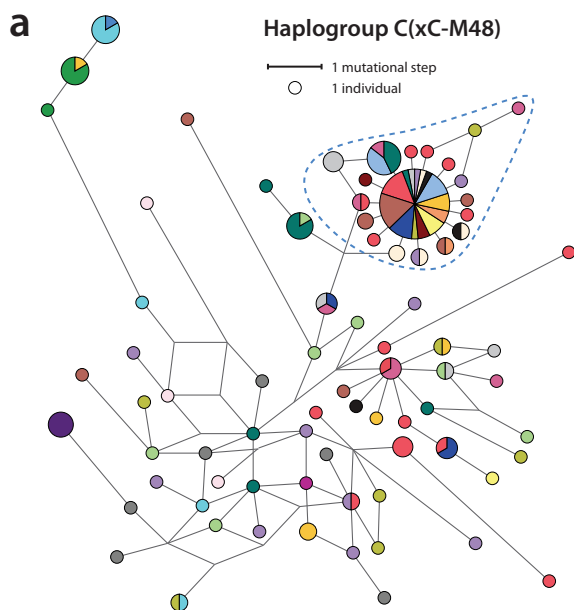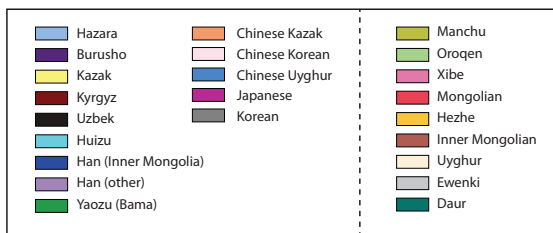Brown GR, Laland KN, Mulder MB (2009) Bateman's principles and human sex roles. Trends Ecol Evol 24: 297-304.

Caballero A (1995) On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139: 1007-11.

Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, Bradley DG (2015) Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proc Natl Acad Sci U S A.

Chiaroni J, Underhill PA, Cavalli-Sforza LL (2009) Y chromosome diversity, human expansion, drift, and cultural evolution. Proc Natl Acad Sci U S A 106: 20174-9.

Creanza N, Feldman MW (2016) Worldwide genetic and cultural change in human evolution. Curr Opin Genet Dev 41: 85-92.

de Knijff P, Kayser M, Caglià A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold G, Roewer L (1997) Chromosome Y microsatellites: population genetic and evolutionary aspects. Int. J. Legal Med. 110: 134-140.

Francalacci P, Morelli L, Angius A, Berutti R, Reinier F, Atzeni R, Pilu R, Busonero F, Maschio A, Zara I, Sanna D, Useli A, Urru MF, Marcelli M, Cusano R, Oppo M, Zoledziewska M, Pitzalis M, Deidda F, Porcu E, Poddie F, Kang HM, Lyons R, Tarrier B, Gresham JB, Li B, Tofanelli S, Alonso S, Dei M, Lai S, Mulas A, Whalen MB, Uzzau S, Jones C, Schlessinger D, Abecasis GR, Sanna S, Sidore C, Cucca F (2013) Low-pass DNA sequencing of 1200 Sardinians reconstructs European Y-chromosome phylogeny. Science 341: 565-9.

Fu Q, Li H, Moorjani P, Jay F, Slepchenko SM, Bondarev AA, Johnson PL, Aximu-Petri A, Prufer K, de Filippo C, Meyer M, Zwyns N, Salazar-Garcia DC, Kuzmin YV, Keates SG, Kosintsev PA, Razhev DI, Richards MP, Peristov NV, Lachmann M, Douka K, Higham TF, Slatkin M, Hublin JJ, Reich D, Kelso J, Viola TB, Paabo S (2014) Genome sequence of a 45,000-year-old modern human from western Siberia. Nature 514: 445-9.

Gibbons A (1997) Y chromosome shows that Adam was an African. Science 278: 804-805.

Goldberg A, Gunther T, Rosenberg NA, Jakobsson M (2017) Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proc Natl Acad Sci U S A 114: 2657-2662.

Guillot EG, Cox MP (2015) High Frequency Haplotypes are Expected Events, not Historical Figures. F1000Res 4: 666.

Gusmão L, Sanchez-Diz P, Calafell F, Martin P, Alonso CA, Alvarez-Fernandez F, Alves C, Borjas-Fajardo L, Bozzo WR, Bravo ML, Builes JJ, Capilla J, Carvalho M, Castillo C, Catanesi CI, Corach D, Di Lonardo AM, Espinheira R, Fagundes de Carvalho E, Farfan MJ, Figueiredo HP, Gomes I, Lojo MM, Marino M, Pinheiro MF, Pontes ML, Prieto V, Ramos-Luis E, Riancho JA, Souza Goes AC, Santapa OA, Sumita DR, Vallejo G, Vidal Rioja L, Vide MC, Vieira da Silva CI, Whittle MR, Zabala W, Zarrabeitia MT, Alonso A, Carracedo A, Amorim A (2005) Mutation rates at Y chromosome specific microsatellites. Hum. Mutat. 26: 520-528.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney E, Stewardson K, Fu Q, Mittnik A, Banffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Rojo Guerra MA, Roth C, Szecsenyi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, Reich D (2015) Massive migration from the steppe was a source for Indo-European languages in Europe. Nature 522: 207-11.

Hallast P, Batini C, Zadik D, Maisano Delser P, Wetton JH, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Destro Bisol G, Dupuy BM, Eriksen HA, Jorde LB, King TE, Larmuseau MH, Lopez de Munain A, Lopez-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Schempp W, Sears M, Tolun A, Tyler-Smith C, Van Geystelen A, Watkins S, Winney B, Jobling MA (2015)
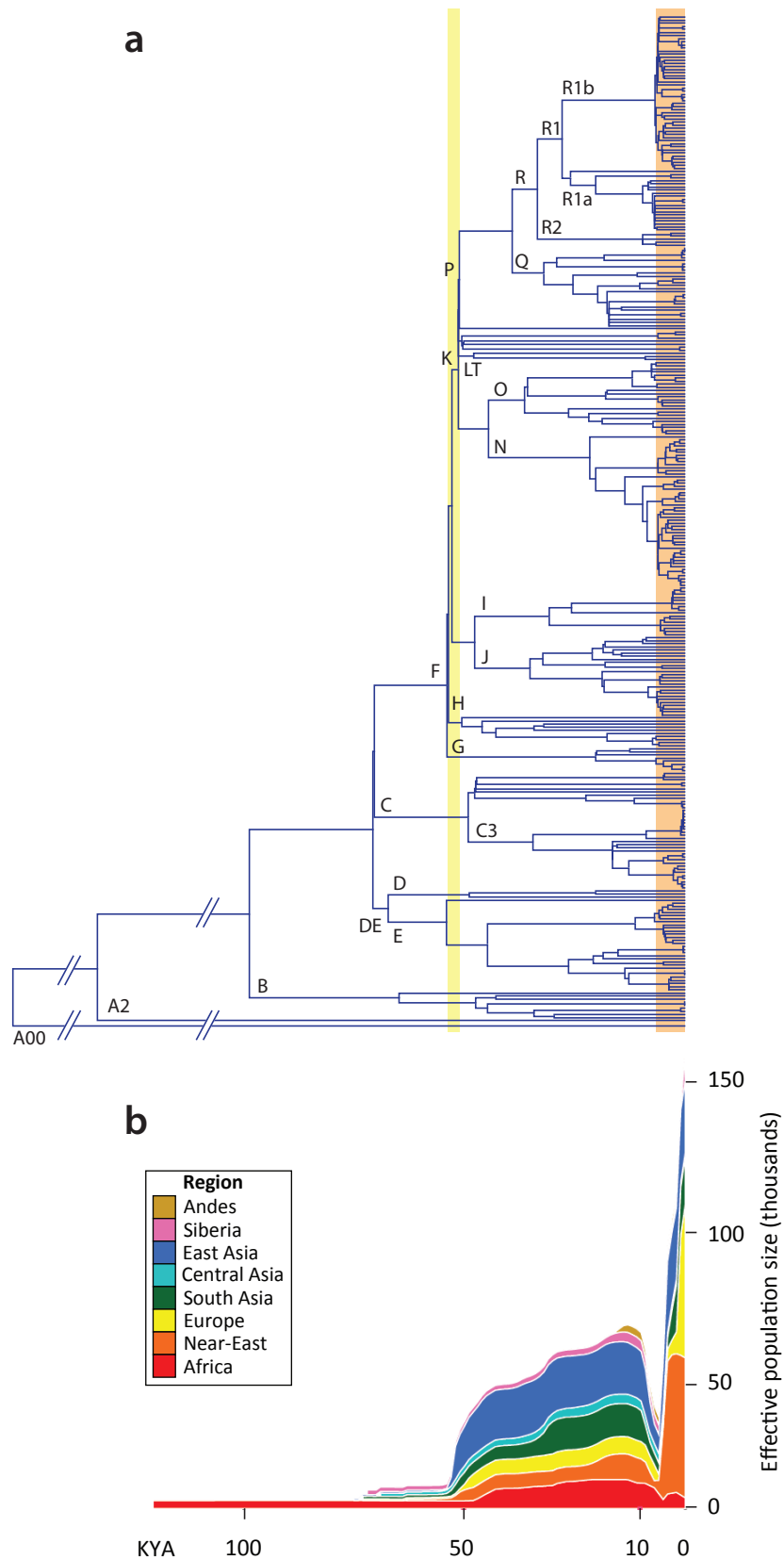
The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. Mol Biol Evol 32: 661-73.

Helgason A, Einarsson AW, Guethmundsdottir VB, Sigurethsson A, Gunnarsdottir ED, Jagadeesan A, Ebenesersdottir SS, Kong A, Stefansson K (2015) The Y-chromosome point mutation rate in humans. Nat Genet 47: 453-7.

Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Graves T, Fulton RS, Dugan S, Ding Y, Buhay CJ, Kremitzki C, Wang Q, Shen H, Holder M, Villasana D, Nazareth LV, Cree A, Courtney L, Veizer J, Kotkiewicz H, Cho TJ, Koutseva N, Rozen S, Muzny DM, Warren WC, Gibbs RA, Wilson RK, Page DC (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. Nature 483: 82-6.

Jobling MA, Hollox EJ, Hurles ME, Kivisild T, Tyler-Smith C (2014) Human Evolutionary Genetics, 2nd edn. Garland Science, New York and London

Jobling MA, Pandya A, Tyler-Smith C (1997) The Y chromosome in forensic analysis and paternity testing. Int. J. Legal Med. 110: 118-124.

Karafet TM, Mendez FL, Meilerman M, Underhill PA, Zegura SL, Hammer MF (2008) New binary polymorphisms reshape and increase resolution of the human Y-chromosomal haplogroup tree. Genome Res. 18: 830-838.

Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, Talas UG, Rootsi S, Ilumae AM, Magi R, Mitt M, Pagani L, Puurand T, Faltyskova Z, Clemente F, Cardona A, Metspalu E, Sahakyan H, Yunusbayev B, Hudjashov G, DeGiorgio M, Loogvali EL, Eichstaedt C, Eelmets M, Chaubey G, Tambets K, Litvinov S, Mormina M, Xue Y, Ayub Q, Zoraqi G, Korneliussen TS, Akhatova F, Lachance J, Tishkoff S, Momynaliev K, Ricaut FX, Kusuma P, Razafindrazaka H, Pierron D, Cox MP, Sultana GN, Willerslev R, Muller C, Westaway M, Lambert D, Skaro V, Kovacevic L, Turdikulova S, Dalimova D, Khusainova R, Trofimova N, Akhmetova V, Khidiyatova I, Lichman DV, Isakova J, Pocheshkhova E, Sabitov Z, Barashkov NA, Nymadawa P, Mihailov E, Seng JW, Evseeva I, Migliano AB, Abdullah S, Andriadze G, Primorac D, Atramentova L, Utevska O, Yepiskoposyan L, Marjanovic D, Kushniarevich A, Behar DM, Gilissen C, Vissers L, Veltman JA, Balanovska E, Derenko M, Malyarchuk B, Metspalu A, Fedorova S, Eriksson A, Manica A, Mendez FL, Karafet TM, Veeramah KR, Bradman N, Hammer MF, Osipova LP, Balanovsky O, Khusnutdinova EK, Johnsen K, Remm M, Thomas MG, Tyler-Smith C, Underhill PA, Willerslev E, Nielsen R, Metspalu M, Villems R, Kivisild T (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res 25: 459-66.

King TE, Jobling MA (2009) Founders, drift and infidelity: the relationship between Y chromosome diversity and patrilineal surnames. Mol. Biol. Evol. 26: 1093-1102.

Lewin R (1987) The unmasking of mitochondrial Eve. Science 238: 24–26.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schroder R, Stoneking M (2014) Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investig Genet 5: 13.

Lkhagvasuren G, Shin H, Lee SE, Tumen D, Kim JH, Kim KY, Kim K, Park AJ, Lee HW, Kim MJ, Choi J, Choi JH, Min NY, Lee KH (2016) Molecular Genealogy of a Mongol Queen's Family and Her Possible Kinship with Genghis Khan. PLoS One 11: e0161622.

Maisano Delser P, Neumann R, Ballereau S, Hallast P, Batini C, Zadik D, Jobling MA (2017) Signatures of human European Palaeolithic expansion shown by resequencing of non-recombining X-chromosome segments. Eur J Hum Genet.

Moore LT, McEvoy B, Cape E, Simms K, Bradley DG (2006) A Y-chromosome signature of hegemony in Gaelic ireland. Am. J. Hum. Genet. 78: 334-338.

Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E (2017) Tracing the peopling of the world through genomics. Nature 541: 302-310.

Pardo LM, Piras G, Asproni R, van der Gaag KJ, Gabbas A, Ruiz-Linares A, de Knijff P, Monne M, Rizzu

P, Heutink P (2012) Dissecting the genetic make-up of North-East Sardinia using a large set of haploid and autosomal markers. Eur J Hum Genet 20: 956-64.

Pinhasi R, Thomas MG, Hofreiter M, Currat M, Burger J (2012) The genetic history of Europeans. Trends Genet 28: 496-505.

Poznik GD, Henn BM, Yee MC, Sliwerska E, Euskirchen GM, Lin AA, Snyder M, Quintana-Murci L, Kidd JM, Underhill PA, Bustamante CD (2013) Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. Science 341: 562-5.

Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Wilson Sayres MA, Ayub Q, McCarthy SA, Narechania A, Kashin S, Chen Y, Banerjee R, Rodriguez-Flores JL, Cerezo M, Shao H, Gymrek M, Malhotra A, Louzada S, Desalle R, Ritchie GR, Cerveira E, Fitzgerald TW, Garrison E, Marcketta A, Mittelman D, Romanovitch M, Zhang C, Zheng-Bradley X, Abecasis GR, McCarroll SA, Flicek P, Underhill PA, Coin L, Zerbino DR, Yang F, Lee C, Clarke L, Auton A, Erlich Y, Handsaker RE, 1000 Genomes Project Consortium, Bustamante CD, Tyler-Smith C (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet 48: 593-9.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16: 1791-8.

Raghavan M, Steinrucken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Avila-Arcos MC, Malaspinas AS, Eriksson A, Moltke I, Metspalu M, Homburger JR, Wall J, Cornejo OE, Moreno-Mayar JV, Korneliussen TS, Pierre T, Rasmussen M, Campos PF, Damgaard Pde B, Allentoft ME, Lindo J, Metspalu E, Rodriguez-Varela R, Mansilla J, Henrickson C, Seguin-Orlando A, Malmstrom H, Stafford T, Jr., Shringarpure SS, Moreno-Estrada A, Karmin M, Tambets K, Bergstrom A, Xue Y, Warmuth V, Friend AD, Singarayer J, Valdes P, Balloux F, Leboreiro I, Vera JL, Rangel-Villalobos H, Pettener D, Luiselli D, Davis LG, Heyer E, Zollikofer CP, Ponce de Leon MS, Smith CI, Grimes V, Pike KA, Deal M, Fuller BT, Arriaza B, Standen V, Luz MF, Ricaut F, Guidon N, Osipova L, Voevoda MI, Posukh OL, Balanovsky O, Lavryashina M, Bogunov Y, Khusnutdinova E, Gubina M, Balanovska E, Fedorova S, Litvinov S, Malyarchuk B, Derenko M, Mosher MJ, Archer D, Cybulski J, Petzelt B, Mitchell J, Worl R, Norman PJ, Parham P, Kemp BM, Kivisild T, Tyler-Smith C, Sandhu MS, Crawford M, Villems R, Smith DG, Waters MR, Goebel T, Johnson JR, Malhi RS, Jakobsson M, Meltzer DJ, Manica A, Durbin R, Bustamante CD, Song YS, Nielsen R, et al. (2015) Genomic evidence for the Pleistocene and recent population history of Native Americans. Science 349: aab3884.

Rootsi S, Magri C, Kivisild T, Benuzzi G, Help H, Bermisheva M, Kutuev I, Barac L, Pericic M, Balanovsky O, Pshenichnov A, Dion D, Grobei M, Zhivotovsky LA, Battaglia V, Achilli A, Al-Zahery N, Parik J, King R, Cinnioglu C, Khusnutdinova E, Rudan P, Balanovska E, Scheffrahn W, Simonescu M, Brehm A, Goncalves R, Rosa A, Moisan JP, Chaventre A, Ferak V, Furedi S, Oefner PJ, Shen P, Beckman L, Mikerezi I, Terzic R, Primorac D, Cambon-Thomsen A, Krumina A, Torroni A, Underhill PA, Santachiara-Benerecetti AS, Villems R, Semino O (2004) Phylogeography of Y-chromosome haplogroup I reveals distinct domains of prehistoric gene flow in Europe. Am. J. Hum. Genet. 75: 128-137.

Rozen S, Marszalek JD, Alagappan RK, Skaletsky H, Page DC (2009) Remarkably little variation in proteins encoded by the Y chromosome's single-copy genes, implying effective purifying selection. Am J Hum Genet 85: 923-8.

Scozzari R, Massaia A, Trombetta B, Bellusci G, Myres NM, Novelletto A, Cruciani F (2014) An unbiased resource of novel SNP markers provides a new chronology for the human Y chromosome and reveals a deep phylogenetic structure in Africa. Genome Res. 24: 535-544.

Semino O, Passarino G, Oefner PJ, Lin AA, Arbuzova S, Beckman LE, De Benedictis G, Francalacci P, Kouvatsi A, Limborska S, Marcikiae M, Mika A, Mika B, Primorac D, Santachiara-Benerecetti

AS, Cavalli-Sforza LL, Underhill PA (2000) The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: a Y chromosome perspective. Science 290: 1155-1159.

Trombetta B, D'Atanasio E, Massaia A, Ippoliti M, Coppa A, Candilio F, Coia V, Russo G, Dugoujon JM, Moral P, Akar N, Sellitto D, Valesini G, Novelletto A, Scozzari R, Cruciani F (2015) Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent. Genome Biol Evol 7: 1940-50.

Webster TH, Wilson Sayres MA (2016) Genomic signatures of sex-biased demography: progress and prospects. Curr Opin Genet Dev 41: 62-71.

Wei W, Ayub Q, Chen Y, McCarthy S, Hou Y, Carbone I, Xue Y, Tyler-Smith C (2013) A calibrated human Y-chromosomal phylogeny based on resequencing. Genome Res 23: 388-95.

Wilder JA, Mobasher Z, Hammer MF (2004) Genetic evidence for unequal effective population sizes of human females and males. Mol. Biol. Evol.

Willems T, Gymrek M, Poznik GD, Tyler-Smith C, Genomes Project Chromosome YG, Erlich Y (2016) Population-Scale Sequencing Data Enable Precise Estimates of Y-STR Mutation Rates. Am J Hum Genet 98: 919-33.

Wilson Sayres MA, Lohmueller KE, Nielsen R (2014) Natural selection reduced diversity on human Y chromosomes. PLoS Genet 10: e1004064.

Xue Y, Zerjal T, Bao W, Zhu S, Lim SK, Shu Q, Xu J, Du R, Fu S, Li P, Yang H, Tyler-Smith C (2005) Recent spread of a Y-chromosomal lineage in northern China and Mongolia. Am. J. Hum. Genet. 77: 1112-1116.

Yan S, Wang CC, Zheng HX, Wang W, Qin ZD, Wei LH, Wang Y, Pan XD, Fu WQ, He YG, Xiong LJ, Jin WF, Li SL, An Y, Li H, Jin L (2014) Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. PLoS One 9: e105691.

Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, Qamar R, Ayub Q, Mohyuddin A, Fu S, Li P, Yuldasheva N, Ruzibakiev R, Xu J, Shu Q, Du R, Yang H, Hurles ME, Robinson E, Gerelsaikhan T, Dashnyam B, Mehdi SQ, Tyler-Smith C (2003) The genetic legacy of the Mongols. Am. J. Hum. Genet. 72: 717-721.

Zhivotovsky LA, Underhill PA, Cinnioglu C, Kayser M, Morar B, Kivisild T, Scozzari R, Cruciani F, Destro-Bisol G, Spedini G, Chambers GK, Herrera RJ, Yong KK, Gresham D, Tournev I, Feldman MW, Kalaydjieva L (2004) The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. Am. J. Hum. Genet. 74: 50-61.

Legend:
- Hazara
- Burusho
- Kazak
- Kyrgyz
- Uzbek
- Huizu
- Han (Inner Mongolia)
- Han (other)
- Yaozu (Bama)
- Chinese Kazak
- Chinese Korean
- Chinese Uyghur
- Japanese
- Korean
- Manchu
- Oroqen
- Xibe
- Mongolian
- Hezhe
- Inner Mongolian
- Uyghur
- Ewenki
- Daur

a Haplogroup C(xC-M48)
- 1 mutational step
- 1 individual

b Haplogroup C-M48
- 1 mutational step
- 1 individual

15 Y-STRs

61 Y-STRs

c Haplogroup R1b-M269
- Uí Néill cluster
- other
- 1 mutational step
- 1 individual

Batini & Jobling, Figure 1

a

R1b
R1
R
R1a
R2
Q
P
K
LT
O
N
I
J
F
H
G
C
C3
D
DE
E
B
A2
A00

b

Region
Andes
Siberia
East Asia
Central Asia
South Asia
Europe
Near-East
Africa

Effective population size (thousands)

150

100

50

0

KYA   100   50   10   0

Batini & Jobling, Figure 2 (revised)