International Congress Series 1261 (2004) 70–72

www.ics-elsevier.com

The distribution of Y-chromosomal haplotypes: forensic implications

Mark A. Jobling*, Turi E. King

Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Abstract. Over 200 single nucleotide polymorphisms (SNPs) and ~40 short tandem repeats (STRs) make the Y chromosome the most informative haplotyping system in the genome. The SNPs define haplogroups forming a unique phylogeny, and with high geographical differentiation, leading to interest in predicting population-of-origin. Data on admixed populations suggest that such predictions may often be misleading. Studies of relationships between surnames and Y haplotypes suggest that surname prediction from DNA will not be reliable. © 2003 Elsevier B.V. All rights reserved.

Keywords: Admixture; Haplogroup; Population genetics; Surnames; Y chromosome

1. Introduction

The human Y chromosome is sex-determining, and therefore male-specific. The non-recombining region thus represents a useful source of polymorphisms for the forensic analysis of male DNA. Since most criminals are men, Y markers are potentially widely applicable. In fact, powerful autosomal short tandem repeat (STR) multiplexes deal effectively with most situations, and Y-STRs find use in specialized applications, including the analysis of mixed samples in rape cases. Here, we address the issues of haplotype distribution, prediction of population-of-origin, and the relationship between Y haplotypes and surnames.

2. Y-SNPs and haplogroups

In evolutionary studies, the most useful Y polymorphisms are binary markers, mostly single nucleotide polymorphisms (SNPs), of which over 200 are currently well characterized. Because mutation rates are low ($\sim 10^{-8}$ per SNP per generation), the haplotypes these SNPs define ('haplogroups'), form a unique bifurcating phylogeny. In an important advance, haplogroup nomenclature has recently been standardized [1].

Abbreviations: SNP, single nucleotide polymorphism; STR, short tandem repeat.

* Corresponding author. Tel.: +44-1162-523377; fax: +44-1162-523378.
E-mail address: maj4@leicester.ac.uk (M.A. Jobling).

doi:10.1016/S0531-5131(03)01787-4

Though this tree represents the best-resolved phylogeny of any human locus, problems remain. Many of the markers have been discovered in a limited sample set, then are typed more widely. This can elevate the apparent diversity in the populations used for discovery—ascertainment bias. Many branches in the phylogeny bear multiple markers, but usually a single marker is typed, so intermediate haplogroups, important for refining tree resolution, will not be found. Typing methods are nonstandardized and control DNAs from particular haplogroups are usually unavailable. Finally, there is no centralized database of haplogroup data—researchers must trawl the literature and often decipher nonstandard nomenclature.

More SNPs are required to resolve parts of the tree and it is tempting to seek these in existing SNP databases. However, ascertainment bias is serious here, since a small number of populations are represented in screening sets, and there is also the problem of validation: many apparent Y-SNPs are actually sequence differences between highly similar duplicated sequences (paralogs) that have been misassembled.

3. Haplogroup distributions

Global surveys reveal a striking picture of geographical haplogroup differentiation [2], and many haplogroups are continent-specific. This specificity is greater that that seen with autosomal or mitochondrial DNA (mtDNA) markers and is explained by the prevalence of patrilocality, a marriage practice where men stay closer to their birthplaces upon marriage than do women [3]. Can we therefore use the Y chromosome to predict population-of-origin of a DNA sample? Studies of admixed populations caution against a simplistic approach. In a sample of British Afro-Caribbeans, all but $\sim$ 1% of individuals carry African mtDNAs, but >25% of Y chromosomes are of European origin (unpublished observations). This reflects sex-biased admixture, where European men, but not women, contributed genes to the ancestors of these people during the period of slavery. Admixture is often sex-biased, and therefore the Y is relatively sensitive to introgression, which could mislead if Y haplogroup were used to guide an investigation. If good population data were available, a likelihood could be presented of population-of-origin of a DNA sample, and this might still constitute useful information ; however, useful markers are probably best found elsewhere in the genome.

4. Y-STRs present and future

Around 30 Y-STRs are in current use and 9 are employed in excellent quality-controlled databases [4]; haplotype diversity is high (virtual heterozygosity = 0.9976 in the European database). Each haplogroup originates as a new SNP mutation in a single man and he carries a single Y-STR haplotype. As time passes, STR mutations accumulate in his descendants and haplotype diversity increases. Eventual diversity depends on time and demography, but is reduced compared to that of Y chromosomes as a whole. So, just as haplogroups show geographical specificity, so do Y-STR haplotypes, though rarely in such a clear way. High-frequency Y-STR haplotypes are not usually encountered, but can be where a single man and his descendants have been particularly successful in reproducing. A good example is the presence of a frequent (8% of sample) 16-locus haplotype over an enormous geographical area stretching from the Pacific to the Caspian Sea and attributed to the activities of Genghis Khan [5].

Known STRs are a mixed bag of different markers with different properties. Availability of the Y chromosome sequence, and the program Tandem Repeats Finder [6], allows systematic searches for further useful STRs. These have yielded well over 100 markers (Kayser et al., in preparation), which will allow STRs with particular properties to be used in the future.

5. Y chromosomes and surnames

The fact that many societies employ patrilineal surnames has stimulated interest in the relationship between surname and Y haplotype. In a perfect world, a detailed Y haplotype would automatically provide the surname of the bearer. However, in reality, there are a several perturbing influences: some names had more than one founder (about 25 generations ago in England); non-paternities and adoptions have introduced other lineages; and people sometimes change their surnames [7]. The only published study, of the surname Sykes [8], suggested either multiple founders or a high non-paternity rate. Even with a rate as low as 1%, the cumulative rate over the ~ 50 generations separating two contemporary men from a shared surname founder is $\sim 40\%$, so we expect to see multiple lineages in most surnames. In our current studies, most of a set of 40 different surnames show multiple Y-STR clusters, each of which represents either an independent surname founder, or, more probably, an independent non-paternity event. Simulation studies are being used to choose between different explanations, but it seems that surname prediction from Y haplotype is likely to remain a fantasy.

Acknowledgements

MAJ was supported by a Wellcome Trust Senior Fellowship (grant no. 057559) and TEK by the Wellcome Trust.

References

[1] Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups, Genome Res. 12 (2002) 339–348.
[2] M.A. Jobling, C. Tyler-Smith, The human Y chromosome: an evolutionary marker comes of age, Nat. Rev., Genet. 4 (2003) 598–612.
[3] M.T. Seielstad, et al., Genetic evidence for a higher female migration rate in humans, Nat. Genet. 20 (1998) 278–280.
[4] L. Roewer, et al., Online reference database of Y-chromosomal short tandem repeat (STR) haplotypes, Forensic Sci. Int. 118 (2001) 103–111.
[5] T. Zerjal, et al., The genetic legacy of the Mongols, Am. J. Hum. Genet. 72 (2003) 717–721.
[6] G. Benson, Tandem repeats finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (1999) 573–580.
[7] M.A. Jobling, In the name of the father: surnames and genetics, Trends Genet. 17 (2001) 353–357.
[8] B. Sykes, C. Irven, Surnames and the Y chromosome, Am. J. Hum. Genet. 66 (2000) 1417–1419.