The history of the Y chromosome in man

Jennifer F Hughes & David C Page

Studies of the Y chromosome over the past few decades have opened a window into the history of our species, through the reconstruction and exploitation of a patrilineal (Y-genealogical) tree based on several hundred single-nucleotide variants (SNVs). A new study validates, refines and extends this tree by incorporating >65,000 Y-linked variants identified in 1,244 men representing worldwide diversity.

Despite its puny size, the Y chromosome is among the most storied of human chromosomes. The refined genealogical tree of modern human Y chromosomes reported by Chris Tyler-Smith, Carlos Bustamante and colleagues in this issue[1] extends this narrative legacy and ensures that it will continue well into the future.

Why the Y?

The Y chromosome is home to the testis-determining gene *SRY*, which causes fetuses to develop as anatomical males. The Y chromosome and its meiotic partner, the X chromosome, thereby qualify as sex chromosomes. But here the ironies begin. Across 95% of its length (all but its pseudoautosomal tips), the Y chromosome abstains from sexual recombination—the exchange of genes—with the X chromosome. This has been true throughout the history (and long before the origin) of our species. Thus, the great bulk of the Y chromosome has been transmitted clonally—that is, asexually—from father to son down the generations, its content shaped only by mutation and selection, without swapping genetic material with a partner during meiotic cell division, as occurs on all other nuclear chromosomes, including the X chromosome. The isolationist behavior of the Y chromosome unfortunately resulted in its decline, both in size and in gene repertoire, but it also turned the Y chromosome into a powerful device for recording the migratory and demographic history of the males of our species. Random, stable mutations, such as SNVs, accumulate over time across the Y chromosome, and particular collections of these mutations are permanently linked together because of the Y chromosome's asexual transmission. A set of human males who share a particular collection of Y-chromosome mutations is called a haplogroup, and those shared mutations can be traced back to a common patrilineal ancestor. Y-chromosome transmission across generations is analogous to an (asexually) expanding yeast colony, where random mutations accumulate in some cells and are passed on to progeny during cell division. Some such mutations in yeast have striking phenotypic consequences, and their origin and spread can be traced through the colony (**Fig. 1**). In the same way, Y-chromosome SNV patterns gathered from men across the world have been used to delineate a detailed phylogenetic tree showing relationships among extant Y chromosomes, which gives insight into our species' history[2–4].

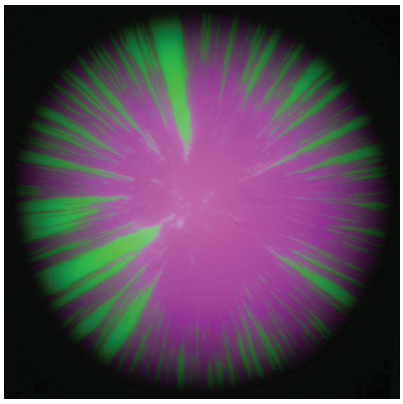
Jennifer F. Hughes is at the Whitehead Institute, Cambridge, Massachusetts, USA, and David C. Page is at the Whitehead Institute, Howard Hughes Medical Institute, and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.
e-mail: jhughes@wi.mit.edu

Tracing human history

The new study by Poznik *et al.*[1] uses the unprecedented data set of whole-genome sequences generated by the 1000 Genomes Project[5], which includes individuals selected from 26 geographically diverse human populations. In total, the authors analyzed Y-chromosome sequences from 1,244 men. They restricted their analysis to single-copy regions of the Y chromosome, which equate ~10 million base pairs, or roughly half of the Y chromosome's total euchromatic sequence. This strategy ensured that the sequence variants they identified represented polymorphisms (differences between orthologous sequences in individuals) rather than differences between paralogous copies of long (>10,000-base-pair), nearly identical repeats, which are prominent on the Y chromosome. Because the Y chromosome is present in only one copy per cell, the authors had to contend with relatively low sequence coverage (~4.3-fold). The authors were therefore conservative in their approach, and their account of >65,000 variants (including SNVs, insertions, deletions and short-tandem-repeat length variation) is likely an underestimate of the true level of variation among the Y chromosomes studied.

On the basis of the distribution of SNVs, the authors constructed a phylogenetic tree showing the paternal lineages connecting all of the 1,244 men studied. The structure of this new tree mirrors previous trees, which were based on 100-fold fewer SNVs[3,4,6]. However, the wealth of information provided by the large number of both Y chromosomes and variants included in the current study further refines the tree and provides new insight into human population dynamics[1]. For example, the expanded analysis identifies a new mega-group, which originated ~55,000 years ago and encompasses nearly all non-African males, and an ancient clade within haplogroup H, which is prominent in South Asia. At least eight major population expansions are evident in the new phylogenetic tree, and the timing of a number of these expansions correlates with notable events in human history.

Unlike earlier studies that focused on a set of previously ascertained SNVs, the new study is unbiased because it catalogs all variation detected across a large expanse of the Y chromosome. This approach enables a robust method for calculating dates of bifurcations, or population splits, within the tree. The authors estimate the date for the most recent common ancestor of all Y chromosomes, which is rooted in Africa, to be ~190,000 years and the date for all non-African Y chromosomes to be ~76,000 years. These dates are considerably



Anne E. Dodson and Jasper Rine, 2015

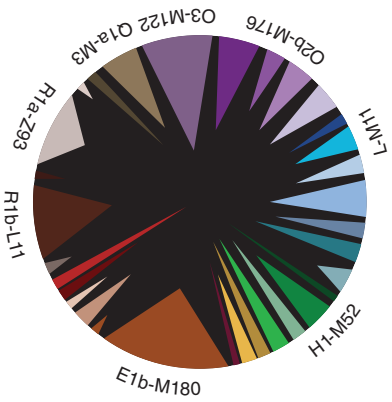


Figure 1 Clonal expansion of mutants in yeast and man. Clonal expansion of mutants in a yeast colony (left) and among a geographically diverse collection of human Y chromosomes (right; an artist's reinterpretation of Fig. 2 from Poznik *et al.*[1]). Pink-to-green mutants in the yeast propagate as clonal sectors.

older than those reported in previous but less expansive Y-chromosome sequencing efforts[7–9]; the discrepancies are due in part to differences in the Y-chromosome mutation rate used for calibration.

Importantly, this new data set provides abundant information for conducting future studies of Y-chromosome genealogy. However, tracking the inheritance of the Y chromosome provides a masculinized, and in some instances misleading, view of the history of our species. For example, reproductive success can differ dramatically among individuals because of societal or political factors, which could lead to the predominance of one particular Y chromosome in a population. This observation of reduced Y-chromosome diversity could be misinterpreted as evidence of a severe population bottleneck. The most famous example of this phenomenon is seen in the legacy of Genghis Khan, whose Y-chromosome descendants comprise ~8% of men in a large region of Asia[10], and another study found additional examples of the spread of such 'super Ys' in recent human history[11]. The Y chromosome provides just one piece of a complex puzzle of human history, with mitochondrial DNA, which is maternally inherited, and autosomal loci filling in the rest. Just as this present study was not possible a decade ago, ten years from now, we will likely have effectively unlimited access to genome-wide variation from across the globe, allowing a more encompassing reconstruction of our species' history.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Poznik, G.D. *et al. Nat. Genet.* **48**, 593–599 (2016).
2. Underhill, P.A. *et al. Nat. Genet.* **26**, 358–361 (2000).
3. Y Chromosome Consortium. *Genome Res.* **12**, 339–348 (2002).
4. Karafet, T.M. *et al. Genome Res.* **18**, 830–838 (2008).
5. 1000 Genomes Project Consortium. *Nature* **526**, 68–74 (2015).
6. Jobling, M. & Tyler-Smith, C. *Nat. Rev. Genet.* **4**, 598–612 (2003).
7. Hallast, P. *et al. Mol. Biol. Evol.* **32**, 661–673 (2015).
8. Poznik, G.D. *et al. Science* **341**, 562–565 (2013).
9. Wei, W. *et al. Genome Res.* **23**, 388–395 (2013).
10. Zerjal, T. *et al. Am. J. Hum. Genet.* **72**, 717–721 (2003).
11. Balaresque, P. *et al. Eur. J. Hum. Genet.* **23**, 1413–1422 (2015).

The carrot genome sequence brings colors out of the dark

Jordi Garcia-Mas & Manuel Rodriguez-Concepcion

The genome sequence of carrot (*Daucus carota* L.) is the first completed for an Apiaceae species, furthering knowledge of the evolution of the important euasterid II clade. Analyzing the whole-genome sequence allowed for the identification of a gene that may regulate the accumulation of carotenoids in the root.

Carrot is among the ten most important vegetable crops and shows a worldwide distribution[1]. Carrots (that is, the storage roots of *D. carota* plants) are well known for their nutritional value due to the high content of provitamin carotenoids (alpha- and beta-carotene)[2]. Carotenoids also provide the orange, yellow and red colors characteristic of these vegetables. Indeed, the name of these pigments derives from *carota*, the Latin word for carrot, and was coined in 1831 by the German pharmacist H.W.F. Wackenroder[3]. Carrot belongs to the Apiaceae (Umbelliferae) family, which contains about 3,700 species, including celery, parsley, parsnip, and many herb and spice plants. Carrot was domesticated in Central Asia as a root crop around 1,100 years ago, and its distribution then expanded to both the east and west[4]. Whereas wild carrots are white and produce low levels of carotenoids, the first domesticated carrots were purple and yellow because of the accumulation of anthocyanins and the carotenoid lutein, respectively. Orange carrots accumulating alpha- and beta-carotene were not reported until the sixteenth century in Europe, and red carrots, which contain the red carotenoid lycopene, were seen in Asia in the eighteenth century. Modern carrot varietal groups differ in root color, shape and size, some of the chief traits in carrot breeding programs. On page 657 of this issue, Philipp Simon and colleagues[5] report a high-quality assembly of the carrot genome and propose a mechanism to explain what makes carrot so unique among plants: the production and accumulation of astounding levels of carotenoids in the storage root.

Iorizzo *et al.* sequenced the genome of a double-haploid Nantes-type orange carrot using Illumina and BAC end sequences, reporting a genome assembly representing 89.1% of the 473-Mb estimated genome size. The assembly statistics suggest that this is a high-quality genome, comparable to other recently reported plant genomes of similar size. The carrot genome assembly consists of 46% repetitive sequences and 32,113 gene models. Iorizzo *et al.* also resequenced 35 accessions representing the existing variability in carrot, including wild and cultivated accessions of both eastern and western origin. The observed phylogeny of these accessions is in agreement with placing the primary center of carrot domestication in Central Asia[4]. Iorizzo *et al.* describe several chromosomal regions showing

Jordi Garcia-Mas and Manuel Rodriguez-Concepcion are at the Center for Research in Agricultural Genomics (CRAG) CSIC-IRTA-UAB-UB, Barcelona, Spain, and Jordi Garcia-Mas is also at the Institut de Recerca i Tecnologia Agroalimentàries, Barcelona, Spain.
e-mail: jordi.garcia@irta.cat