

## Report

### The Genetic Legacy of the Mongols

Tatiana Zerjal,<sup>1</sup> Yali Xue,<sup>1,2</sup> Giorgio Bertorelle,<sup>3</sup> R. Spencer Wells,<sup>4</sup> Weidong Bao,<sup>1,5</sup> Suling Zhu,<sup>1,5</sup> Raheel Qamar,<sup>1,6</sup> Qasim Ayub,<sup>1,6</sup> Aisha Mohyuddin,<sup>1,6</sup> Songbin Fu,<sup>2</sup> Pu Li,<sup>2</sup> Nadira Yuldasheva,<sup>4,7</sup> Ruslan Ruzibakiev,<sup>7</sup> Jiujin Xu,<sup>5</sup> Qunfang Shu,<sup>5</sup> Ruofu Du,<sup>5</sup> Huanming Yang,<sup>5</sup> Matthew E. Hurles,<sup>8</sup> Elizabeth Robinson,<sup>1,\*</sup> Tudevdayva Gerelsaikhan,<sup>1,†</sup> Bumbein Dashnyam,<sup>9</sup> S. Qasim Mehdi,<sup>5</sup> and Chris Tyler-Smith<sup>1</sup>

<sup>1</sup>Department of Biochemistry, University of Oxford, Oxford; <sup>2</sup>Department of Medical Biology, Harbin Medical University, Harbin, China; <sup>3</sup>Dipartimento di Biologia, Università di Ferrara, Ferrara, Italy; <sup>4</sup>Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, United Kingdom; <sup>5</sup>Institute of Genetics, Chinese Academy of Sciences, Beijing; <sup>6</sup>Biomedical and Genetic Engineering Labs, Islamabad; <sup>7</sup>Institute of Immunology, Academy of Sciences, Tashkent, Uzbekistan; <sup>8</sup>McDonald Institute, University of Cambridge, Cambridge, United Kingdom; and <sup>9</sup>Institute of Biotechnology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia

**We have identified a Y-chromosomal lineage with several unusual features. It was found in 16 populations throughout a large region of Asia, stretching from the Pacific to the Caspian Sea, and was present at high frequency: ~8% of the men in this region carry it, and it thus makes up ~0.5% of the world total. The pattern of variation within the lineage suggested that it originated in Mongolia ~1,000 years ago. Such a rapid spread cannot have occurred by chance; it must have been a result of selection. The lineage is carried by likely male-line descendants of Genghis Khan, and we therefore propose that it has spread by a novel form of social selection resulting from their behavior.**

The patterns of variation found in human DNA are usually considered to result from a balance between neutral processes and natural selection. Among the former, mutation, recombination, and migration increase variation, whereas genetic drift decreases it. Natural selection can act to remove deleterious variants (purifying selection), maintain polymorphism (balancing selection), or produce a trend (directional selection). Clear examples of the latter are rare in humans, but probable cases, such as those associated with resistance to malaria (Hamblin and Di Rienzo 2000) or unidentified pathogens (Stephens et al. 1998), can be recognized by the “signature” they leave in the genome. The rapid increase in frequency of the selected allele and its linked sequences results in a haplotype that is found at higher frequency than would be expected from its degree of variation. We have now

identified such a haplotype on the Y chromosome, but we suggest that its spread results not from a biological advantage, but from human activities recorded in history.

In surveys of DNA variation in Asia, we typed 2,123 men with  $\geq 32$  markers to produce a Y haplotype for each man; these included 1,126 individuals described elsewhere (Qamar et al. 2002; Zerjal et al. 2002). Over 90% of the haplotypes showed the usual pattern (Mohyuddin et al. 2001): most males had a unique code; and the few haplotypes present in more than one individual were generally found within the same population. However, we also saw one pattern that was novel in two respects. First, there was a high frequency of a cluster of closely related lineages, collectively called the “star cluster” (fig. 1, *shaded area*). Second, star-cluster chromosomes were found in 16 populations throughout a large geographical area extending from Central Asia to the Pacific (fig. 2); thus, they do not result from an event specific to any single population. We can deduce the most likely time to the most recent common ancestor (TMRCA) and place of origin of this unusual lineage from the observed genetic variation. To do this, it is first necessary to distinguish star-cluster chromosomes from the remainder. For this, we used the criterion that haplotypes linked to the central one in the shaded area of the network without gaps would be included (fig. 1).

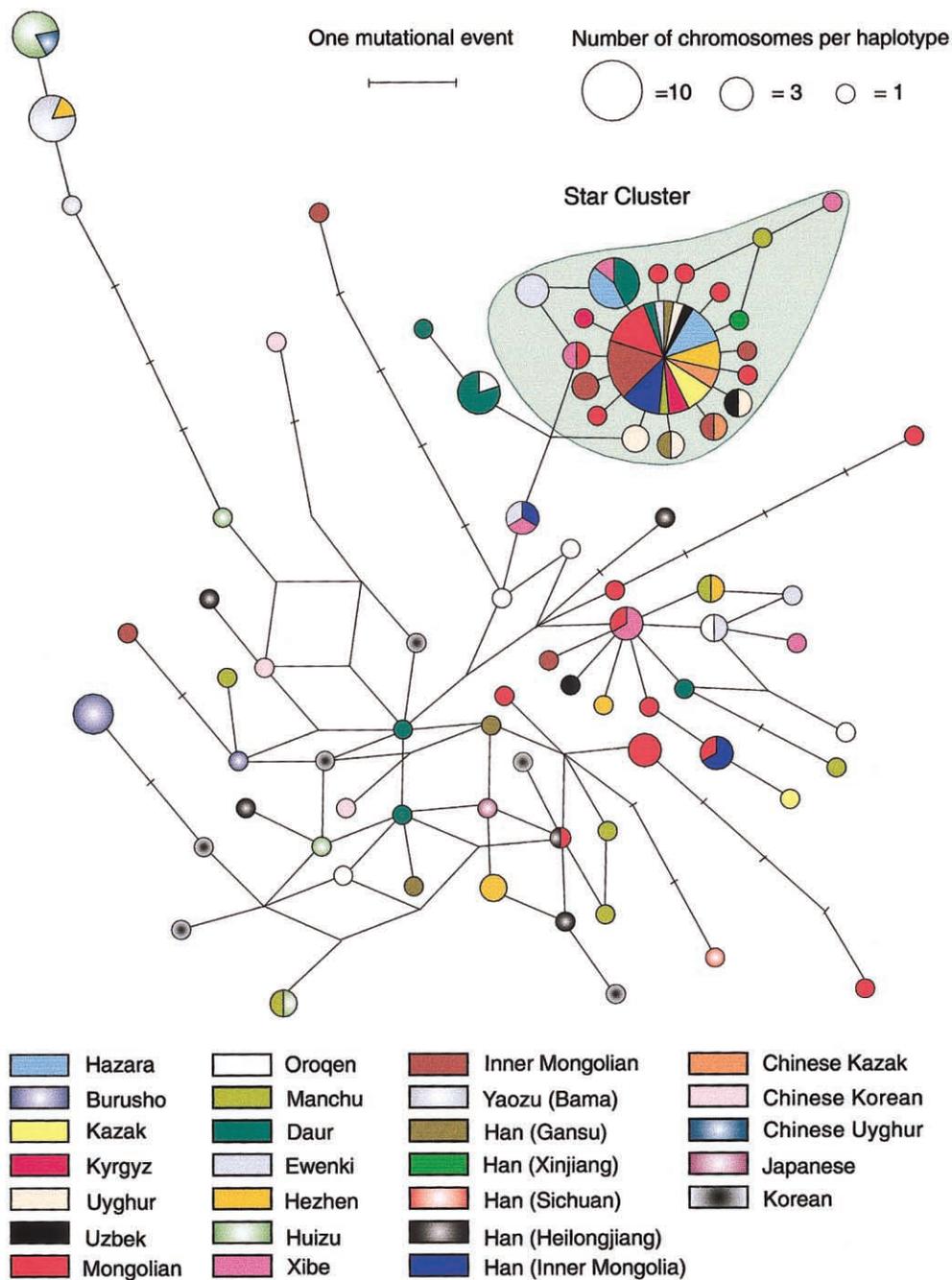
Received September 27, 2002; accepted for publication November 25, 2002; electronically published January 17, 2003.

Address for correspondence and reprints: Dr. Chris Tyler-Smith, Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. E-mail: chris@bioch.ox.ac.uk

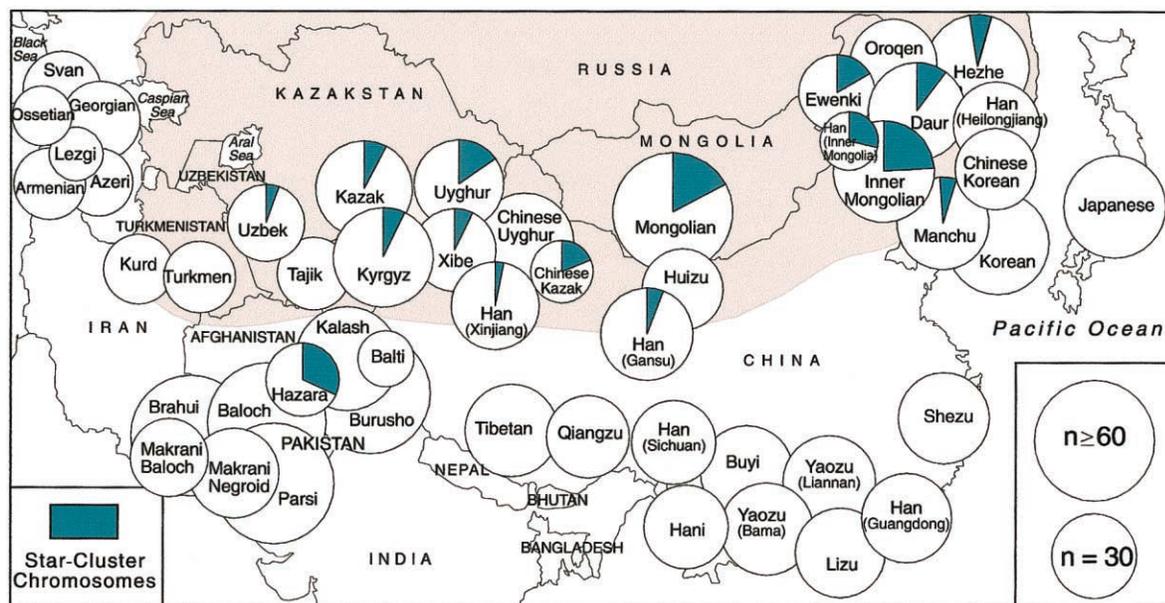
\* Present affiliation: Combinatorx, Boston

† Present affiliation: National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda.

© 2003 by The American Society of Human Genetics. All rights reserved. 0002-9297/2003/7203-0022\$15.00



**Figure 1** Median-joining network (Bandelt et al. 1999) representing Y-chromosomal variation within haplogroup C\*(x3c3c). Chromosomes were typed with a minimum of 16 binary markers (Qamar et al. 2002; Zerjal et al. 2002; our unpublished observations), including RPS4Y and M48, to define the lineage C\*(x3c3c) (Y-Chromosome Consortium 2002), also known as haplogroup 10, derived for RPS4Y and ancestral for C. Sixteen Y microsatellites were also typed, but DYS19 was excluded from the network analysis because it is duplicated in haplogroup C. The central star-cluster profile is 10-16-25-10-11-13-14-12-11-11-11-12-8-10-10, for the loci DYS389I-DYS389b-DYS390-DYS391-DYS392-DYS393-DYS388-DYS425-DYS426-DYS434-DYS435-DYS436-DYS437-DYS438-DYS439. Circles represent lineages, area is proportional to frequency, and color indicates population of origin. Lines represent microsatellite mutational differences.



**Figure 2** Geographical distribution of star-cluster chromosomes. Populations are shown as circles with an area proportional to sample size; star-cluster chromosomes are indicated by green sectors. The shaded area represents the extent of Genghis Khan’s empire at the time of his death (Morgan 1986).

We then used two approaches to calculate a TMRCA for the star-cluster chromosomes. The program BAT-WING (Wilson and Balding 1998) uses models of both mutation and population processes, which were specified as described elsewhere (Qamar et al. 2002). With this program, we estimated ~1,000 years for the TMRCA (95% confidence interval limits ~700–1,300 years). The use of alternative demographic models with constant or exponentially increasing population size changed the estimate by <10%. A method that does not consider population structure (Morral et al. 1994),  $\rho$ , suggested ~860 (~590–1,300) years. In both calculations, we assumed a generation time of 30 years. The origin was most likely in Mongolia, where the largest number of different star-cluster haplotypes is found (fig. 1). Thus, a single male line, probably originating in Mongolia, has spread in the last ~1,000 years to represent ~8% of the males in a region stretching from northeast China to Uzbekistan. If this spread were due to a general population expansion, we would expect to find multiple lineages with the same characteristics of high frequency and presence in multiple populations, but we do not (Zerjal et al. 2002). The star-cluster pattern is unique.

This rise in frequency, if spread evenly over ~34 generations, would require an average increase by a factor of ~1.36 per generation and is thus comparable to the most extreme selective events observed in natural populations, such as the spread of melanic moths in 19th-century England in response to industrial pollution

(Edleston 1865). We evaluated whether it could have occurred by chance. If the population growth rate is known, it is possible to test whether the observed frequency of a lineage is consistent with its level of variation, assuming neutrality (Slatkin and Bertorelle 2001). Using this method, we estimated the chance of finding the low degree of variation observed in the star cluster, with a current frequency of ~8%, under neutral conditions. Even with the demographic model most likely to lead to rapid increase of the lineage, double exponential growth, the probability was  $<10^{-237}$ ; if the mutation rate were 10 times lower, the probability would still be  $<10^{-10}$ . Thus, chance can be excluded: selection must have acted on this haplotype.

Could biological selection be responsible? Although this possibility cannot be entirely ruled out, the small number of genes on the Y chromosome and their specialized functions provide few opportunities for selection (Jobling and Tyler-Smith 2000). It is therefore necessary to look for alternative explanations. Increased reproductive fitness, transmitted socially from generation to generation, of males carrying the same Y chromosome would lead to the increase in frequency of their Y lineage, and this effect would be enhanced by the elimination of unrelated males. Within the last 1,000 years in this part of the world, these conditions are met by Genghis (Chingis) Khan (c. 1162–1227) and his male relatives. He established the largest land empire in history and often slaughtered the conquered populations, and he and his

close male relatives had many children. Although the Mongol empire soon disintegrated as a political unit, his male-line descendants ruled large areas of Asia for many generations. These included China, where the Yüan Dynasty emperors remained in power until 1368, after which the Mongols continued to dominate the country north of the Great Wall for several more centuries, and the region west to the Aral Sea, where the Chaghatai Khans ruled. Although their power diminished over time, they remained at Kashghar near the Kyrgyzstan/China border until the middle of the 17th century (Morgan 1986).

It is striking that the boundary of the Mongol empire when Genghis Khan died (fig. 2), which also corresponds to the boundaries of the regions controlled by later Khans, matches the distribution of star-cluster chromosomes closely, with one exception: the Hazaras. We, therefore, wished to compare Genghis Khan's Y profile with the star cluster. It is not possible to examine his remains directly, but history provides an alternative. The Hazaras of Pakistan have a Mongol origin (Qamar et al. 2002), and many consider themselves to be direct male-line descendants of Genghis Khan. A genealogy documenting these links has been constructed from their oral history (Mousavi 1998). A large proportion of the Hazara profiles do indeed lie in the star cluster, which is not otherwise seen in Pakistan (fig. 2), thus supporting their oral tradition and suggesting that Genghis Khan carried the star-cluster haplotype.

The Y chromosome of a single individual has spread rapidly and is now found in ~8% of the males throughout a large part of Asia. Indeed, if our sample is representative, this chromosome will be present in about 16 million men, ~0.5% of the world's total. The available evidence suggests that it was carried by Genghis Khan. His Y chromosome would obviously have had ancestors, and our best estimate of the TMRCA of star-cluster chromosomes lies several generations before his birth. Several scenarios, which are not mutually exclusive, could explain its rapid spread: (1) all populations carrying star-cluster chromosomes could have descended from a common ancestral population in which it was present at high frequency; (2) many or most Mongols at the time of the Mongol empire could have carried these chromosomes; (3) it could have been restricted to Genghis Khan and his close male-line relatives, and this specific lineage could have spread as a result of their activities. Explanation 1 is unlikely because these populations do not share other Y haplotypes, and explanation 2 is difficult to reconcile with the high Y-haplotype diversity of modern Mongolians (Zerjal et al. 2002). The historically documented events accompanying the establishment of the Mongol empire would have contributed directly to the spread of this lineage by Genghis

Khan and his relatives, but perhaps as important was the establishment of a long-lasting male dynasty. This scenario shows selection acting on a group of related men; group selection has been much discussed (Wilson and Sober 1994) and is distinguished by the property that the increased fitness of the group is not reducible to the increased fitness of the individuals. It is unclear whether this is the case here. Our findings nevertheless demonstrate a novel form of selection in human populations on the basis of social prestige. A founder effect of this magnitude will have influenced allele frequencies elsewhere in the genome: mitochondrial DNA lineages will not be affected, since males do not transmit their mitochondrial DNA, but, in the simplest models, the founder male will have been the ancestor of each autosomal sequence in ~4% of the population and X-chromosomal sequence in ~2.7%, with implications for the medical genetics of the region. Large-scale changes to patterns of human genetic variation can occur very quickly. Although local influences of this kind may have been common in human populations, it is, perhaps, fortunate that events of this magnitude have been rare.

## Acknowledgments

We thank all DNA donors, for making this work possible; Ed Southern, for encouragement and comments on the manuscript; Jaume Bertranpetit, for helpful discussions; Ian Wilson, for modifying BATWING to allow 16 different mutation rates to be used; Askar Mousavi, for information on the Hazaras; Christine Keyser, for information about the genetics of ancient Mongolian populations; and Felipe Fernández-Armesto, for historical discussions. This work was supported by a Joint Project from the Royal Society and the National Natural Science Foundation of China and a Collaborative Research Initiative Grant from the Wellcome Trust; T. Z. was supported by the Wellcome Trust.

## References

- Bandelt HJ, Forster P, Rohl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37–48
- Edleston RS (1865) *Amphydasis betularia*. *Entomologist* 2:150
- Hamblin MT, Di Rienzo A (2000) Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *Am J Hum Genet* 66:1669–1679
- Jobling MA, Tyler-Smith C (2000) New uses for new haplotypes: the human Y chromosome, disease and selection. *Trends Genet* 16:356–362
- Mohyuddin A, Ayub Q, Qamar R, Zerjal T, Helgason A, Mehdi SQ, Tyler-Smith C (2001) Y-chromosomal STR haplotypes in Pakistani populations. *Forensic Sci Int* 118:141–146
- Morgan D (1986) *The Mongols*. Blackwell Publishers, Oxford
- Morral N, Bertranpetit J, Estivill X, Nunes V, Casals T, Gi-

- menez J, Reis A, et al (1994) The origin of the major cystic fibrosis mutation ( $\Delta F508$ ) in European populations. *Nat Genet* 7:169–175
- Mousavi SA (1998) *The Hazaras of Afghanistan*. Curzon Press, Richmond
- Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, Zerjal T, Tyler-Smith C, Mehdi SQ (2002) Y-chromosomal DNA variation in Pakistan. *Am J Hum Genet* 70:1107–1124
- Slatkin M, Bertorelle G (2001) The use of intraallelic variability for testing neutrality and estimating population growth rate. *Genetics* 158:865–874
- Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, et al (1998) Dating the origin of the CCR5- $\Delta 32$  AIDS-resistance allele by the coalescence of haplotypes. *Am J Hum Genet* 62:1507–1515
- Wilson DS, Sober S (1994) Re-introducing group selection to the human behavioral sciences. *Behavioral Brain Sci* 17:585–654
- Wilson IJ, Balding DJ (1998) Genealogical inference from microsatellite data. *Genetics* 150:499–510
- Y-Chromosome-Consortium (2002) A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res* 12:339–348
- Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002) A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am J Hum Genet* 71:466–482