THE HUMAN Y CHROMOSOME: AN EVOLUTIONARY MARKER COMES OF AGE

Mark A. Jobling and Chris Tyler-Smith‡*

Until recently, the Y chromosome seemed to fulfil the role of juvenile delinquent among human chromosomes — rich in junk, poor in useful attributes, reluctant to socialize with its neighbours and with an inescapable tendency to degenerate. The availability of the near-complete chromosome sequence, plus many new polymorphisms, a highly resolved phylogeny and insights into its mutation processes, now provide new avenues for investigating human evolution. Y-chromosome research is growing up.

SATELLITE DNA
A large tandemly-repeated DNA array that spans hundreds of kilobases to megabases.

RECOMBINATION
The formation of a new combination of alleles through meiotic crossing over. Some authors include intrachromosomal gene conversion under this heading. As this has been shown on the Y chromosome, they prefer not to refer to it as 'non-recombining'.

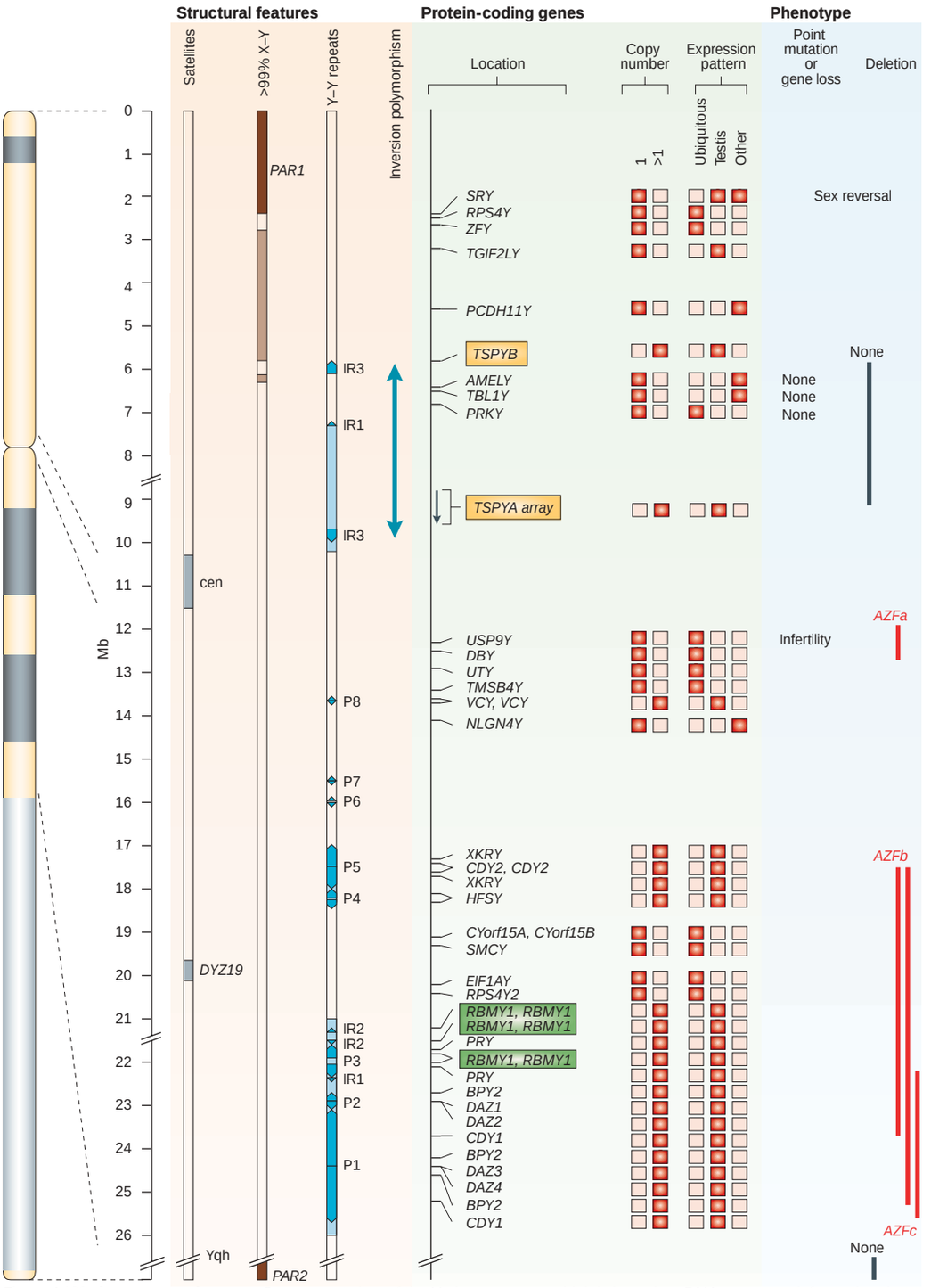
EUCHROMATIN
The part of the genome that is decondensed during interphase, which is transcriptionally active.

**Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK.*
‡The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.
e-mails: maj4@le.ac.uk; cts@sanger.ac.uk
doi:10.1038/nrg1124

The properties of the Y chromosome read like a list of violations of the rulebook of human genetics: it is not essential for the life of an individual (males have it, but females do well without it), one-half consists of tandemly repeated SATELLITE DNA and the rest carries few genes, and most of it does not recombine. However, it is because of this disregard for the rules that the Y chromosome is such a superb tool for investigating recent human evolution from a male perspective and has specialized, but important, roles in medical and forensic genetics.

The Y chromosome needs to be studied in different ways from the rest of the genome. In the absence of RECOMBINATION, genetic mapping provides no information and physical analyses are of paramount importance, so the availability of the near-complete sequence of the EUCHROMATIC portions of the chromosome might have even greater impact here than on the rest of the genome. In this review, we are mainly concerned with the application of Y-chromosomal DNA variation to investigations of human evolution, but there is also considerable overlap with other areas. Questions about phenotype are relevant because phenotypic effects can lead to evolutionary changes being influenced by natural selection rather than by gene flow and genetic drift.

It seems that the more we know about the Y chromosome the more questions we have. Its sequence[1] allows us to get a comprehensive picture of its structure and organization, and to compile a catalogue of genes (FIG. 1): but how complete is this catalogue, what proportion of genes can be identified and what do they do? Efforts to discover genome-wide sequence variation have identified vast numbers of Y-specific single nucleotide polymorphisms (SNPs): the Ensembl database lists 28,650 at the time of writing, which might seem enough to provide an extremely detailed PHYLOGENETIC TREE of Y-chromosomal lineages. But how many of these SNPs are real, and how many are artefacts that are produced by unknowingly comparing true Y-chromosomal sequences with similar sequences (PARALOGUES) elsewhere on the same or other chromosomes[2]? Also, are these SNPs a representative set of sequence variants from the human population as a whole? The answer is no, because of ascertainment bias (BOX 1) in the range of populations that were surveyed for variation. As well as this possible treasure trove of unverified SNPs, the availability of Y-chromosome sequence means that there are now more than 200 binary polymorphisms that are well characterized, and 100–200 potentially useful new MICROSATELLITES, as well as the ~30 published polymorphic tri-, tetra- and pentanucleotide repeat markers. Finally, there is a robust and developing phylogeny of Y-chromosomal haplotypes (FIG. 3) that are defined by binary polymorphisms (haplogroups), and a unified nomenclature system[3] that allows diversity data from different research groups to be readily integrated.



PHYLOGENETIC TREE
A diagram that represents the evolutionary relationships between a set of taxa (lineages).

PARALOGUES
Sequences, or genes, that have originated from a common ancestral sequence, or gene, by a duplication event.

MICROSATELLITE
A class of repetitive DNA sequences that are made up of tandemly organized repeats that are 2-8 nucleotides in length. They can be highly polymorphic and are frequently used as molecular markers in population genetics studies.

HAPLOGROUP
A haplotype that is defined by binary markers, which is more stable but less detailed than one defined by microsatellites.

Figure 1 | **Y-chromosomal genes.** Our present view of the Y chromosome, which is based on DNA sequence information that was derived largely from a single HAPLOGROUP-R individual[1]. From left to right: cytogenetic features of the chromosome and their approximate locations, which are numbered from the Yp telomere. Structural features include three satellite regions (cen, *DYZ19* and Yqh), segments of X–Y identity (*PAR1* and *PAR2*; dark brown) and high similarity (mid brown), and Y–Y repeated sequences in which the regions with greatest sequence identity are designated 'IR' for 'inverted repeat' and 'P' for 'palindrome'. An inversion polymorphism on Yp that distinguishes haplogroup P from most other lineages is indicated. The locations of the 27 distinct Y-specific protein-coding genes are shown; some are present in more than one copy and their expression patterns are summarized. Pseudoautosomal genes and Y-specific non-coding transcripts are not shown. On the right, the phenotypes that are associated with gene inactivation or loss are indicated; some deletions produce no detectable phenotype[111] (black) and represent polymorphisms in the population, whereas others result in infertility (*AZFa, AZFb* and *AZFc*) (red), although the contributions of the individual deleted genes are unclear.

Analysis of the near-complete euchromatic sequence of the Y chromosome has called into question many of the cliches that were previously associated with this chromosome[1]. Also, this new knowledge of DNA sequence, genes, polymorphisms and phylogeny, provides a starting point for a new generation of studies into Y-chromosome diversity in human populations, mutation processes and the role of the Y chromosome in disease. In this review, we briefly cover mutation processes and disease studies, and point the reader to more detailed coverage in the bibliography, but largely focus on the use of the Y chromosome as a marker in studies of human evolution.

Special features of the Y chromosome

Why focus on a piece of the genome that only tells us about one-half of the population? Because of its sex-determining role, the Y chromosome is male specific and constitutively haploid. It passes from father to son, and, unlike other chromosomes, largely escapes meiotic recombination (BOX 2). Two segments (the pseudoautosomal regions) do recombine with the X, but these amount to less than 3 Mb of its ~60-Mb length; for the purposes of this review, 'Y chromosome' refers to the non-recombining majority (also variously known as NRY, NRPY and MSY). The importance of escaping recombination is that haplotypes, which are the combinations of allelic states of markers along the chromosome, usually pass intact from generation to generation. They change only by mutation, rather than the more complex reshuffling that other chromosomes experience, and so preserve a simpler record of their history. Using binary polymorphisms with low mutation rates, such as SNPs, a unique phylogeny can therefore easily be constructed.

Assuming a 1:1 sex ratio, the human population can be represented in microcosm by one man and one woman. This couple carry four copies of each autosome and three X chromosomes, but only one Y chromosome. In the population as a whole, the EFFECTIVE POPULATION SIZE of the Y chromosome is therefore expected to be one-quarter of that of any autosome, one-third of that of the X chromosome and similar to that of the effectively haploid mitochondrial DNA (mtDNA). Assuming that the same mutation processes act on all chromosomes, we therefore expect lower sequence diversity on the Y chromosome than elsewhere in the nuclear genome, which is indeed observed[4,5]. We also expect it to be more susceptible to genetic drift, which involves random changes in the frequency of haplotypes owing to sampling from one generation to the next. Drift accelerates the differentiation between groups of Y chromosomes in different populations — a useful property for investigating past events. However, because of drift, the frequencies of haplotypes can change rapidly through time, so, quantifying aspects of these events, such as admixture proportions between populations or past demographic changes, might be unreliable.

Geographical clustering is further influenced by the behaviour of men, who are the bearers of Y chromosomes. Approximately 70% of modern societies practice patrilocality[6,7]: if a man and a woman marry but are not from the same place, it is the woman who moves rather than the man. As a consequence, most men live closer to their birthplaces than do women, and local differentiation of Y chromosomes is enhanced. By contrast, mtDNA, which is transmitted only by women, is expected to show reduced geographical clustering. This has been shown in studies of Europe[8] and island Southeast Asia[9], and also in comparisons of populations that practice patrilocality or matrilocality (in which men move and women do not) in Thailand[10]; here, matrilocal groups show enhanced mtDNA differentiation and reduced differentiation of Y chromosomes.

Mutation and Y-chromosome diversity

As mutation is the only force that acts to diversify Y haplotypes, understanding mutational dynamics is important to understanding the origins of haplotype diversity. More generally, knowledge of the rates and processes of mutation at different classes of marker is fundamental to evolutionary interpretations of diversity data and to understanding genetic disease. In this broader context, studies on the Y chromosome are of particular interest because the mutations observed here are the result of exclusively intra-allelic processes. Although these processes occur on all chromosomes, the haploid Y provides a model for studying them without the complicating factors of inter-allelic events and allelic diversity.

NRY, NRPY AND MSY
Several neologisms have been introduced to refer to the portion of the Y chromosome that excludes the pseudoautosomal regions, for example, non-recombining Y (NRY), non-recombining portion Y (NRPY) and male-specific Y (MSY), but none has achieved wide acceptance.

EFFECTIVE POPULATION SIZE
The size of an idealized population that shows the same amount of genetic drift as the population studied. This is approximately 10,000 individuals for humans, in contrast to the census population size of $>6 \times 10^9$.

Box 1 | Ascertainment bias: a bugbear of human evolutionary studies

The best way to discover the single nucleotide polymorphism (SNP) variation in a population is to resequence the chosen region in all individuals. This approach is standard for the highly variable control region of mitochondrial DNA (mtDNA), and has been used in some studies of X-chromosomal and autosomal segments; however, it is rarely used on the Y chromosome because of low variation and high cost. Most published surveys of Y variation have analysed variants that were previously discovered in a different (often small) set of chromosomes. This can cause ascertainment bias, which is systematic distortion in a data set that is caused by the way in which markers or samples are collected. FIG. 2 shows how misleading this can be. If, for example, variants were discovered in European Y chromosomes and then tested in China, the erroneous conclusion would be drawn that Chinese Y chromosomes showed little variation, and *vice versa*. A good illustration of the contradictory conclusions that can be reached is provided by two studies of Y-chromosome diversity in China. Using SNPs that were chosen because of their high variability in the south, Su *et al.*[91] concluded that there was less variation in the north, where only a subset of the southern haplotypes was found. But, when variation was measured by Karafet *et al.* using a more global set of binary markers, more haplotypes were found in the north than in the south[92]. Y-SNP diversity needs to be interpreted with great caution; microsatellites, which are variable in all populations, provide a less biased measure of diversity.

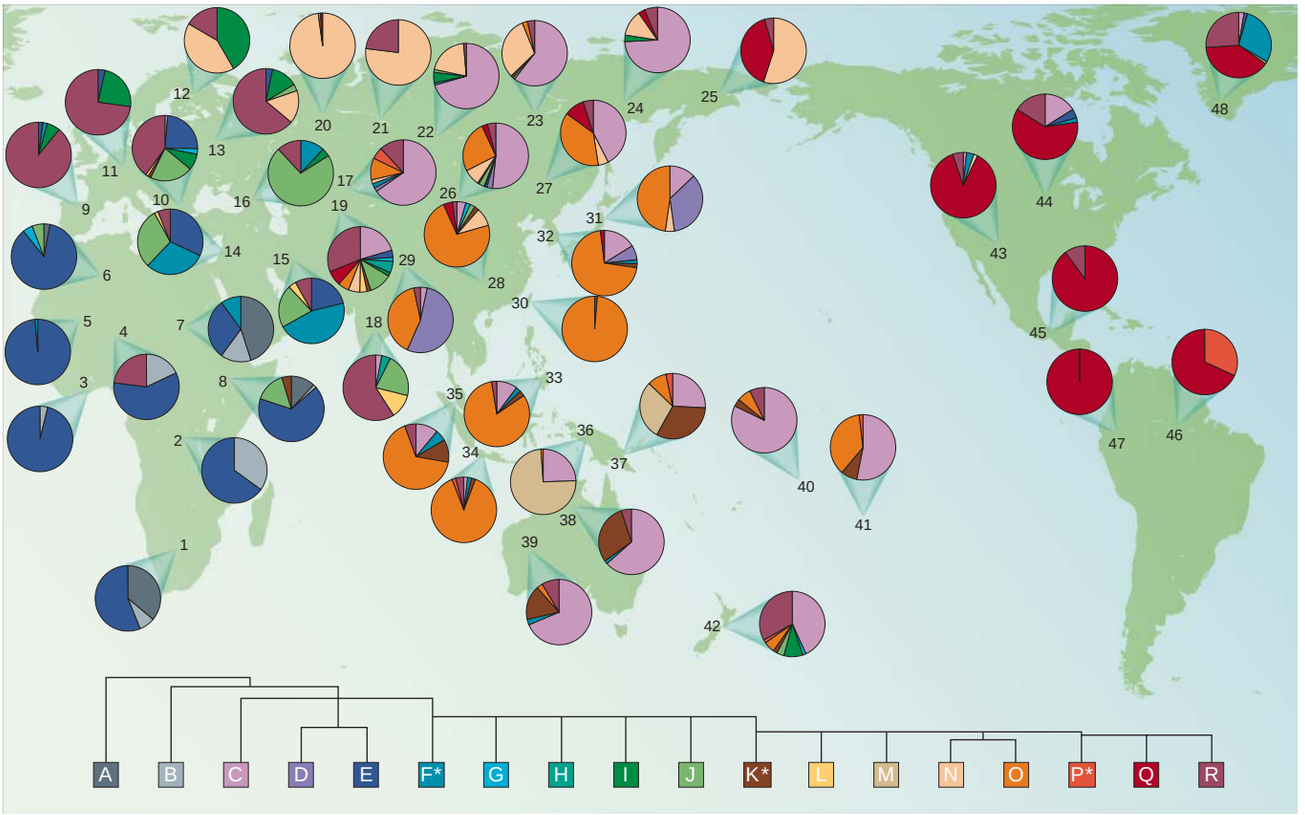


Figure 2 | **Global distribution of Y haplogroups.** Each circle represents a population sample with the frequency of the 18 main Y haplogroups (FIG. 3; shown here in simplified form) identified by the Y Chromosome Consortium (YCC) indicated by the coloured sectors. Note the general similarities between neighbouring populations but large differences between different parts of the world. Data are from published sources (as detailed below) that were chosen to maximize the geographical coverage, sample size and inclusion of relevant markers. In some cases, the limited number of markers could not identify all potential YCC haplogroups and the published data were supplemented by unpublished information from the authors or other studies. Despite this, some populations remain incompletely characterized and the chromosomes assigned here to the PARAGROUPS F*, K* and P*, might be reassigned if further typing is carried out. Populations are numbered as follows: 1, !Kung[60]; 2, Biaka Pygmies[60]; 3, Bamileke[60]; 4, Fali[60]; 5, Senegalese[112]; 6, Berbers[60]; 7, Ethiopians[112]; 8, Sudanese[61]; 9, Basques[61]; 10, Greeks[69]; 11, Polish[69]; 12, Saami[69]; 13, Russians[73]; 14, Lebanese[113]; 15, Iranians[113]; 16, Kazbegi (Georgia)[113]; 17, Kazaks[113]; 18, Punjabis[72]; 19, Uzbeks[73]; 20, Forest Nentsi[73]; 21, Khants[73]; 22, Eastern Evenks[73]; 23, Buryats[73]; 24, Evens[73]; 25, Eskimos[73]; 26, Mongolians[73]; 27, Evenks[73]; 28, Northern Han[73]; 29, Tibetans[78,114]; 30, Taiwanese[61]; 31, Japanese[61]; 32, Koreans[113]; 33, Filipinos[115]; 34, Javanese[115]; 35, Malaysians[115]; 36, West New Guineans (highlands)[115]; 37, Papua New Guineans (coast)[115]; 38, Australians (Arnhem)[115]; 39, Australians (Sandy Desert)[115]; 40, Cook Islanders[115]; 41, Tahitians[116]; 42, Maori[117]; 43, Navajos[78]; 44, Cheyenne[78]; 45, Mixtecs[78]; 46, Makiritare[79]; 47, Cayapa[118]; 48, Greenland Inuit[83].

As with all regions of human DNA — except for the mtDNA control region — base substitutional mutation occurs at too low a rate to be analysed directly. However, the secure phylogenetic framework and haploidy of the Y chromosome mean that recurrent mutations can be identified unambiguously, and data that accumulate from resequencing will provide information about the mutational properties of individual bases. The human population is so large that, even given the low average mutation rate of $\sim 2 \times 10^{-8}$ per base per generation[11], we expect recurrent mutations to occur at every base of the Y chromosome in each global generation. However, these modern recurrences will usually go undetected. At present, the number of recurrent base substitutions in the Y Chromosome Consortium (YCC) tree is only five (REF. 3), although this is likely to increase as more chromosomes are typed.

Studies of genetic diseases show a strong bias towards fathers as the source of new mutations, and also show increasing mutation rate with paternal age (reviewed in REF. 12). The explanations generally used for these two observations are, respectively, the larger number of cell divisions (and hence DNA replications) in male than in female gametogenesis, and the increase of mutation rate with time through continuing divisions of spermatogenic stem cells. As Y chromosomes pass only through the male germline, its mutagenic properties affect the Y chromosome more than any other. The ratio of male to female mutation rates — the α-factor (α) — can be estimated by comparing the number of mutations that have accumulated in homologous autosomal Y-chromosomal and X-chromosomal sequences over a given time period. Estimates for α vary considerably between studies, but all show a significantly higher mutation rate in the male

PARAGROUP
A group of haplotypes that contain some, but not all, of the descendants of an ancestral lineage.

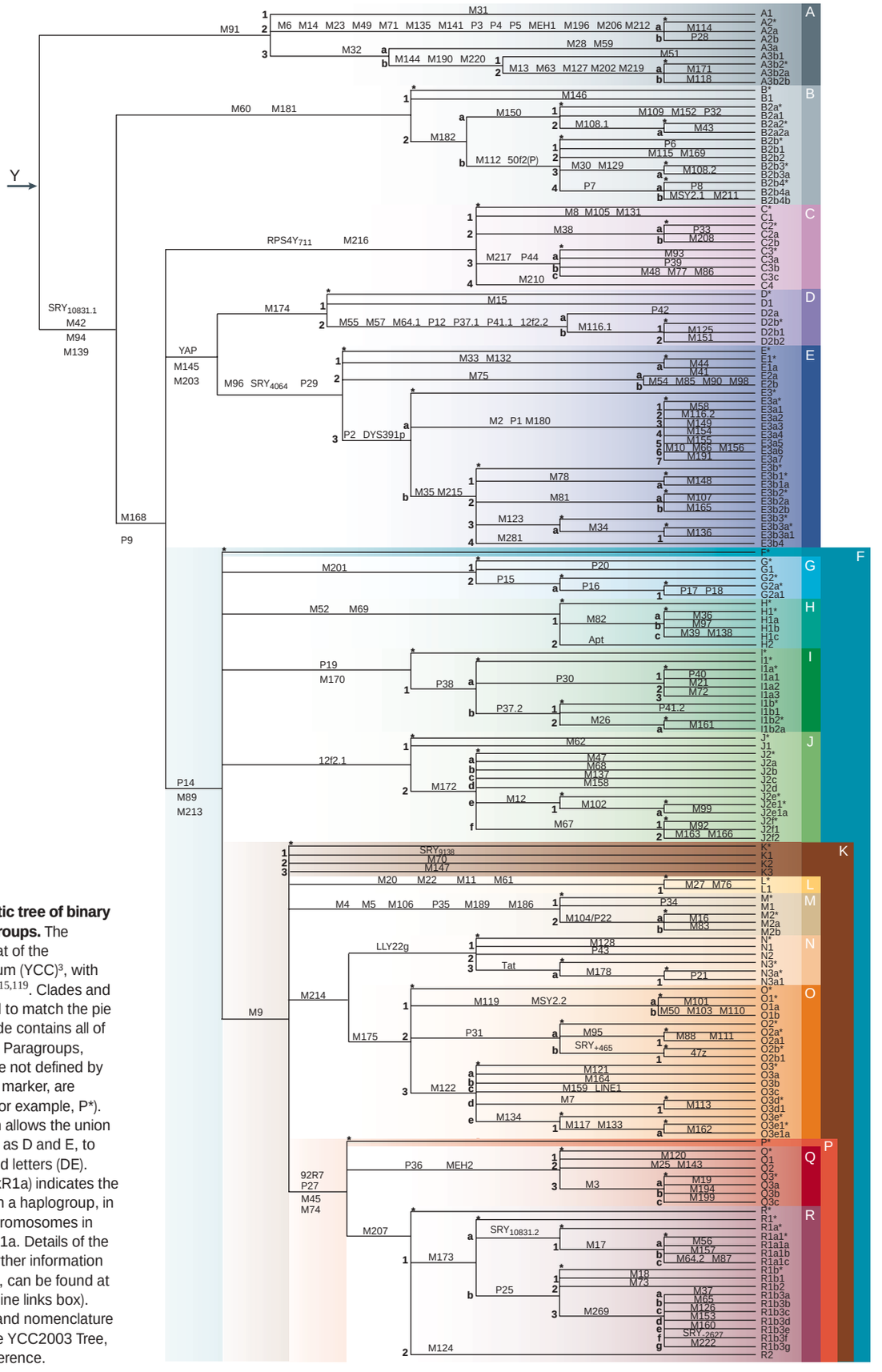


Figure 3 | **The phylogenetic tree of binary Y-chromosomal haplogroups.** The phylogeny is based on that of the Y Chromosome Consortium (YCC)[3], with minor modifications[72,112,115,119]. Clades and clade names are coloured to match the pie charts in FIG. 2. The Y clade contains all of the the haplogroups A–R. Paragroups, which are lineages that are not defined by the presence of a derived marker, are indicated by an asterisk (for example, P*). The nomenclature system allows the union of two haplogroups, such as D and E, to be indicated by juxtaposed letters (DE). A designation such as R(xR1a) indicates the partial typing of markers in a haplogroup, in this case describing all chromosomes in clade R except those in R1a. Details of the markers, together with further information about nomenclature rules, can be found at the YCC web site (see online links box). Note that this phylogeny and nomenclature should be regarded as the YCC2003 Tree, and may be used as a reference.

Box 2 | Could the Y chromosome recombine?

There have been controversies over whether mitochondrial DNA (mtDNA) recombines[93]; although the Y chromosome (excluding the pseudoautosomal regions) is assumed to be exempt from recombination, are there any circumstances under which it could occur? Recent work[20] has shown extensive gene conversion between paralogues on the Y chromosome, and, as gene conversion represents a form of recombination, the argument has been made that the Y chromosome is therefore a recombining chromosome. However, here we adhere to the conventional definition of recombination as crossing over that occurs between chromosomal homologues, and therefore exclude gene conversion from this discussion.

Approximately 1 in 1,000 men[94] carry two Y chromosomes (47,XYY), and, in principle, Y–Y recombination might occur here. As both chromosomes are identical, there would be no consequences for haplotype integrity, but the dynamics of intrachromosomal recombination between repeated sequences could alter. In practice, however, most fertile XYY men eliminate one Y chromosome in their germline[95], so there is no possibility of recombination. Those that retain all three sex chromosomes tend to suffer spermatogenic arrest[96], so any potential recombinants never enter the next generation and are not observed.

Recombination might occur between highly similar XY-HOMOLOGOUS sequences and introduce a block of X-chromosomal material into a Y chromosome. This should be recognizable if the transferred segment is not too small, as the minimum sequence divergence between X and Y chromosomes in such blocks (~1%) is much greater than the average sequence divergence between different Y haplotypes (~0.05%). The rate of such events is probably low: they were observed only twice in an interspecific phylogeny of the cat family[97].

There are rare instances in which segments of the Y chromosome are carried on other chromosomes as asymptomatic translocations[98,99]. Recombination (or gene conversion) could occur between a segment from one haplogroup and a normal Y chromosome from a different haplogroup. Typing polymorphisms on a resulting recombinant chromosome would show a burst of HOMOPLASY in the phylogeny, corresponding to a set of markers located together in the recombinant segment. As population sample sizes and numbers of markers increase, it seems probable that such rare recombinant chromosomes will eventually be discovered.

than in the female germline. A recent estimate[13], based on a comparison of Y and X sequences, gives a value of 2.8 (95% confidence interval limits: 2.3–3.4).

The Y chromosome is rich in low-copy-number repeated sequences, and non-allelic homologous recombination between these paralogues causes both non-pathogenic rearrangements and male infertility through *AZFa*[14–16], *AZFb*[17] and *AZFc*[18] deletions (FIG. 1). There is evidence of GENE CONVERSION between paralogues in two of these cases[19,20]. Such conversion events are expected to be frequent compared with base substitutions, and might influence the probability of rearrangements by altering the lengths of tracts of sequence identity. The controversial proposal has been made that the arrangement of genes that are essential in spermatogenesis (such as the *DAZ* genes) in multiple copies in paralogous repeats has evolved to protect them, through beneficial Y–Y gene conversion, against the degeneration that results from haploidy[20]. It might be that the lack of crossing over with a homologue makes intrachromosomal exchange more likely in the male-specific region of the Y chromosome than on other chromosomes. Gene conversion is a fundamental and poorly understood process that seems certain to be further clarified by studies on the Y chromosome, within the framework of the phylogeny.

Mutation at Y-specific tri- and tetranucleotide microsatellites has been analysed in deep-rooting pedigrees[21] and father–son pairs[22]. Such experiments typically show few mutation events, and for several markers no mutations at all. Studies that use mutation rates in calculations (BOX 3) therefore often quote average rates, such as 3.17×10^{-3} per microsatellite per generation[22], over eight tetranucleotide repeats (95% confidence interval limits: $1.89–4.94 \times 10^{-3}$). However, it is clear from studies of the relative diversity of individual microsatellites in different lineages (for example, REF. 23) that there are marker- and allele-specific differences in mutation rates. In principle, direct analysis of sperm DNA (which is ideal for Y chromosomes) offers access to these rates. The only study to attempt this gave mutation rates for two tetranucleotide microsatellites that were comparable to those obtained by other methods, but for technical reasons only mutations involving repeat gains were analysed[24]. The overall picture presented by these studies is that the properties of Y-specific microsatellites are not detectably different from those of their autosomal counterparts, which indicates that mutation at these markers might be largely or completely intra-allelic, and is consistent with the widely-held view that replication slippage is the mechanism responsible.

By contrast, the fact that the non-recombining region of the Y chromosome is completely devoid of hypervariable GC-rich minisatellites supports the idea that these loci are largely the by-products of recombination activity[25]. The only highly polymorphic Y-specific minisatellite known is an atypical AT-rich locus[26]. Studies of deep-rooting pedigrees[27] and father–son pairs[28] indicate that its high mutation rate might arise from unequal sister-chromatid exchange, or replication slippage that is facilitated by the secondary structure of repeats.

Selection and Y-chromosome diversity

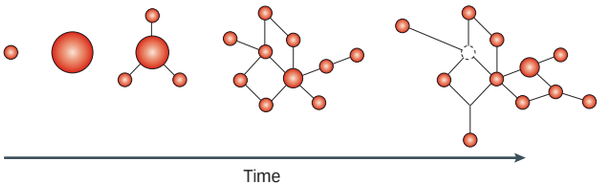
Aside from mutation, selection is a potentially important force in patterning Y-haplotype diversity in populations. The Y chromosome is subject to purifying selection. For example, absence of the Y chromosome (the 45,X karyotype) leads to Turner syndrome, and the loss or inactivation of Y genes can produce an XY female or hermaphrodite phenotype[29] (FIG. 1), or male

HOMOLOGUES
Genes or sequences that share a common ancestor.

HOMOPLASY
The generation of the same sequence state at a locus by independent routes (convergent evolution).

GENE CONVERSION
The non-reciprocal exchange of sequence.

Box 3 | Dating Y lineages using microsatellite variation



A suitable approach to estimate the time to most recent common ancestor (TMRCA) of a set of closely related Y-chromosomes is to use the microsatellite variation in the chromosomes and mutation rates that are derived from family or other studies. The basis of this approach is intuitively obvious: a Y haplogroup originates when a new binary-marker mutation occurs. This happens on a single chromosome, so there is necessarily no associated microsatellite variation. If the lineage spreads, microsatellite mutations will occur: the longer the time, the greater the accumulated variation (see figure). In practice, several factors can complicate the calculation. First, the mutation rate is not known precisely. This is partly because mutations are rare, so the measurement derived from father–son comparisons has a large error, and the use of deep-rooting pedigrees introduces the further problem of distinguishing mutation from non-paternity. However, mutation rates also differ between microsatellites and between alleles (see the section on mutation). Second, a generation time (ranging from 20–35 years) must be chosen to convert a time measured in generations into one measured in years. Third, and perhaps most important, variation will not accumulate at a steady rate, but variants will sometimes be lost by drift, so demography and population structure will have large influences. These are usually poorly known. The effect of these complications is to introduce great uncertainty into the measurements: small errors are more likely to imply that sources of error have been ignored than that a reliable figure has been obtained (see link to BATWING program in the online links box).

infertility[30] (for example, *USP9Y*/*DFFRY*). Here, however, we are concerned with the question of past or present differential selection on Y lineages. BALANCING SELECTION seems unlikely because heterozygous advantage is impossible for this haploid locus, and FREQUENCY-DEPENDENT SELECTION has not been found, but could the chromosome have experienced positive selection when an advantageous mutation arose or a change in circumstances conferred an advantage on a previously neutral variant? Because of the lack of recombination, any selection would affect the entire chromosome and produce an increase in frequency of a lineage more rapidly than would be expected by drift. At first, this lineage would be just one haplogroup among many, and could be detected by comparisons between Y lineages, but eventually it might be fixed and have to be detected by comparisons with other loci.

We expect that polymorphisms in Y-encoded proteins (which are now estimated to number 27 distinct types) might influence some phenotypes (FIG. 1). Association studies, which measure the frequencies of Y haplotypes in matched groups of men with different phenotypes, can detect these influences. However, despite a number of searches over the past five years (TABLE 1), it has been difficult to confirm such effects. Many of the studies showed no association, and in those cases where an association was found, it often could not be reproduced[31,32] or could be explained by population structure[33].

A few associations seem robust. The finding of a higher frequency of XX MALES in haplogroup Y*(xP)[34] can be explained by a plausible mechanism: the protective effect of an inversion polymorphism in haplogroup P, which prevents the ectopic recombination event between the *PRKX* and *PRKY* genes that produces many XX males. However, this will have little evolutionary impact because of the rarity of XX males. The association of low sperm count with haplogroup K(xP) in the Danes deserves further investigation[35]. Because of the high geographical specificity of Y variants, the absence of an association in one population does not imply its absence in other populations in which different lineages predominate, so further studies on a larger scale are warranted to take advantage of improved genotyping methods[36,37] and phylogenetic resolution[3]. Analyses of normal individuals could discover lineages that have expanded more rapidly than expected, irrespective of association to any known phenotype. One such example has been detected, but was explained by social rather than biological selection[38] (BOX 4). So, the evidence available so far does not indicate that significant differential selection acts on contemporary Y lineages.

Under neutrality and the other simplifying assumptions that are commonly made by population geneticists, such as constant population size and random mating, the coalescence time or time to most recent common ancestor (TMRCA) of a locus is proportional to its effective population size. As we have seen, the effective population size of the Y chromosome is one-quarter of that of autosomes and one-third of that of the X chromosome, so a Y-chromosome TMRCA that is significantly less than one-quarter or one-third of these, respectively, would indicate a departure from neutrality, at least as defined by the models. Many estimates of Y-chromosome coalescence time have been made; among those presented in the past few years, some[5,39], but not all[40], are recent. Pritchard *et al.*[39] used data on eight microsatellites from 445 chromosomes to estimate a TMRCA of 46 (16–126) thousand years ago (kilo-years ago, kya) under a model of exponential population growth. Similarly, Thomson *et al.*[5] used SNP variation, which was identified using denaturing high-performance liquid chromatography (DHPLC), in ~66 kb of DNA from 53–70 individuals to infer a TMRCA of 59 (40–140) kya under a similar population model. These point estimates are recent compared with TMRCAs of 177 kya for mtDNA[41], 535 and 1,860 kya for two regions of the X chromosome[42,43] and 800 kya for an autosomal gene[44].

BALANCING SELECTION
Selection that favours more than one allele, for example, through heterozygote advantage, and so maintains polymorphism.

FREQUENCY-DEPENDENT SELECTION
Selection that favours the lower-frequency alleles and so maintains polymorphism.

XX MALE
An individual with a 46,XX karyotype but a male phenotype rather than the expected female phenotype

Table 1 | Association studies using the Y chromosome

Trait	Population	Y marker(s)		Association	Rep.	References
		Binary	*Multi*			
XX maleness	Europeans	2	*MSY1*	XX maleness with haplogroup Y*(xP)	–	34
Infertility/ sperm count	Japanese	3	–	Infertility and low sperm count with haplogroup D	N	31,32
Infertility	Italians	8	–	None	–	33
Infertility	Europeans	11	–	None	–	120
Infertility	Europeans	9	–	None	–	121
Infertility	Japanese	14	αH	None	–	32
Sperm count	Danes	3	–	Low sperm count with haplogroup K(xP)	–	35
Sperm count	Italians	11	–	None	–	122
Alcohol dependence	Finns	1	7	Alcohol dependence with three different lineages	–	123
Height	European-Australians	1	–	Haplogroup P 1.9 cm taller	–	124
Blood pressure	European-Australians	1	–	High blood pressure with haplogroup P	N	125,126
Blood pressure	Polish, Scots	1	–	High blood pressure with haplogroup Y*(xP)	–	125
Blood pressure	Japanese	1	–	None	–	127
Prostate cancer	Japanese	–	1	Prostate cancer with one microsatellite allele	–	128
Testicular cancer	English	7	6	None	–	129
Longevity	Sardinians	14	–	None	–	130
Autism	Norwegians, Swedes, French	10	–	None	–	131

The number of microsatellites or the name of the marker is given in the 'Multi' (multiallelic markers) column. The 'Rep.' column indicates whether the study has been replicated: a blank field indicates that no attempt has been made, whereas 'N' indicates that the association was not replicated.

So, do these relatively recent TMRCA estimates for the Y chromosome provide evidence for a departure from neutrality? This conclusion would be premature: stochastic variation is expected and all estimates have large uncertainties that often do not take into account all sources of error. TMRCAs that are expressed in years require further assumptions about the generation time; this is often set at 20 years with no uncertainty, but, because men on average marry later than women, and because they continue to father children later in life, intergenerational time intervals are longer for male-only lines than for female-only lines. Population genealogy studies give mean intervals of 35 years (men) versus 29 years (women) in Quebec[45], and 31 years (men) versus 28 years (women) in Iceland[46]. The appropriate choice of generation time for ancient populations remains unclear, but the use of 35 years would almost double the TMRCA estimates. Further uncertainty surrounds the chimpanzee–human split (often set at 5 million years ago (mya), which conflicts with fossils that are claimed to be ancestral to humans but not chimpanzees that date from 6–7 mya[47]). Here, we have focused on the implications of TMRCA as evidence for selection. There have been several attempts to apply selection tests that are based on nucleotide diversity (reviewed in REF. 48, and see also REF. 49), which we do not have space to discuss here; however, we note that in these tests, the signal of population expansion is difficult to distinguish from that of selection. For the purposes of this review, we therefore continue to assume that the Y chromosome can be regarded as a neutral locus.

Box 4 | The genetic legacy of Genghis Khan?

Approximately 8% of the chromosomes sampled from a large region of Central Asia (a remarkable ~0.5% of the world total) belong to a closely-related cluster of lineages in haplogroup C with a time to most recent common ancestor (TMRCA) of 1,000 years (95% confidence interval: 700–1,300 years)[38]. Although it is not uncommon for a lineage to drift to predominance in a single small population, this cluster was found in 16 different populations including the Han Chinese, who are the largest ethnic group in the world, and could not have risen to such a high frequency in such a short time by drift alone. The cluster seemed to have originated in Mongolia, and on the basis of its time and place of origin, its geographical distribution (which matched the former Mongol Empire) and its presence in putative male-line descendants of Genghis Khan (circa 1162–1227), the authors suggested that this leader, his male relatives and the dynasty that he founded, were responsible for its spread. The alternative explanation would be that, despite the 20,000 descendants of Genghis Khan reported in 1260, just a century after his birth[100], no trace of his Y chromosome can now be recognized, but that of another man living at the same time in the same place has spread in this unprecedented fashion.

Interpreting diversity patterns

The unmatched phylogenetic and chronological resolution of the Y chromosome (FIG. 3) promises insights into several questions that are of interest to palaeontologists, anthropologists, linguists and historians, and provides the rationale for much of the work involving the Y chromosome: the field of PHYLOGEOGRAPHY. In using the genetic data, we must be wary of simplistic interpretations, such as equating a lineage with a population or a migration, and always remember that we are dealing with one male-specific locus; however, if any single locus can provide useful genetic input into these fields, it is the Y chromosome. Before discussing some of the questions that have been addressed, we need to consider the quality of the information provided by Y-chromosome studies: some aspects are more reliable than others.

The phylogeny is well established. Newly discovered chromosomes and markers are continually refining the tree that is shown in FIG. 3, and it is hoped that the MULTIFURCATIONS will be resolved into BIFURCATIONS, and markers will be found to unify or split unresolved lineages that lack them. It is possible that some important lineages have not yet been sampled, because some areas of the world need more detailed study, but it seems likely that the main branches of the tree are now established.

The chronology is far less certain; as discussed previously, there is debate about the TMRCA of all contemporary Y chromosomes (and corresponding uncertainty about the dates of individual lineages). Young estimates of TMRCA post-date the fossil and archaeological evidence for modern human habitation in geographically separated parts of the world such as Africa and Australia. Under a neutral evolutionary model these estimates are difficult to accept, as they imply recent worldwide replacement of Y chromosomes; so, how should they be interpreted? The confidence intervals that are associated with the TMRCA estimates encompass the earlier dates that are indicated by palaeontology and archaeology. Also, the genetic estimates are based on limited data and demographic models, such as constant-sized or exponentially-increasing populations, which are gross over-simplifications of the complex demographic changes that have occurred during recent human evolution. So, most discussions rely on dates that are determined by palaeontologists and archaeologists, and use the genetic dates only as rough approximations. However, the relative chronology of Y-chromosome coalescence is better established than the absolute chronology, and the dating of recent coalescent events is more reliable than that of older ones. For example, if two Y haplotypes are identical at 100/100 binary markers and 19/20 microsatellites, we can be confident that they shared a common ancestor in the past few generations. By contrast, if they differ at 15/100 binary markers and 14/20 microsatellites, their common ancestor cannot be recent and it is more difficult to make a reliable estimate of when he lived.

Our knowledge of the geographical distribution of each lineage is imprecise. Some populations have not been sampled, and the sample sizes of those investigated are rarely more than a few hundred (often much smaller), so frequencies are known approximately and rare lineages often remain undetected. Evidence of the presence of a lineage is usually reliable, but a lack of evidence does not prove absence. Present distributions are the culmination of many past events. These include some relatively recent ones: intercontinental travel is now common, and will be of great interest to future evolutionary geneticists, but does not concern us today. Similarly, migrations during the past 500 years, although of profound modern epidemiological and forensic significance, are not usually the main focus of attention. It is assumed that recent events can be identified by questioning donors and consulting historical records, and therefore that geographical distributions can provide information about ancient events, but it is important that recent explanations are excluded before ancient ones are sought. For example, when haplogroup Q3 chromosomes, which are characteristic of Native Americans, were detected on the Polynesian island of Rapa[50], did this provide evidence for ancient peopling of Polynesia from America, in accordance with the model of Heyerdahl[51] and in contrast to present models of peopling from Asia? In this case, there was historical evidence for a nineteenth-century slave trade from Peru that was responsible for the Native American presence on Rapa, but if this episode had not been recorded, geneticists could have been misled. The timing of the migration of a lineage is often less clear than in this example.

Overall, the unique features of the Y chromosome combine to produce rapid evolutionary change that is dominated by geographically structured drift. We therefore now review, from a Y-chromosomal perspective, some of the evolutionary questions that have attracted attention. Many of these questions have also been clarified by mtDNA data, and the comparison with Y-chromosome data (although we do not have space to explore it in detail) is often revealing. Further insights promise to emerge from future studies of other loci.

Anthropological insights

Human origins and expansions out of Africa. Members of the genus *Homo* seem to have expanded out of Africa to colonize accessible parts of the rest of the world whenever climatic conditions allowed. There is evidence for *Homo erectus* in Georgia (Caucasus) and Java (southeast Asia) ~1.8–1.6 mya[52,53], whereas subsequent expansions of *Homo heidelbergensis* reached Europe by ~800 kya[54]. Many features of modern anatomy appeared in Africa by 160 kya[55], and anatomically modern human fossils are found in Israel at ~100–90 kya[56], although they were subsequently replaced by Neanderthals[57]. Modern human behaviour, as illustrated by the frequent use of diverse materials for tools, long-distance transport and art, developed even more recently, becoming common only ~70–50 kya in Africa and later in other parts of the world[58]. Other migrations might remain unrecognized, but which events contributed to the gene pool of living humans? Can we trace our ancestry back to several of these ancient populations or were the earlier ones entirely replaced by the later migrants?

PHYLOGEOGRAPHY
The analysis of the geographical distributions of the different branches of a phylogeny.

MULTIFURCATION/ BIFURCATION
The splitting of an ancestral lineage into two or more daughter lineages.

Evidence from the Y chromosome is unequivocal: despite the uncertainty about the Y-chromosome TMRCA discussed earlier, no ancient (>200 kya) Y lineages have been found anywhere in the world, and the Y phylogeny roots in Africa[59], with the first two branches (A and B in FIG. 3) both showing a wide distribution but generally being present at moderate or low frequency (FIG. 2). For example, haplogroup A was found in the !Kung (36%) and Khwe (12%) from the south[60], Malians (2%) from the west[61], and Sudanese (45%) and Ethiopians (14%) from the east[61], but outside Africa only in the Sardinians (5%; 1 individual)[61], which was interpreted as the result of a recent event. Haplogroup B was found in most of the same populations, and also in others including the Biaka (35%) and Mbuti Pygmies (33%), Bamileke (4%) and Fali (18%)[60]. Interestingly, it was also found in two Pakistanis[61] (2%) — does this represent a prehistoric or modern migration? The recent TMRCA, restriction of the most divergent lineages to Africa and evidence for an 'Out-of-Africa' range expansion[59] together show that modern Y diversity arose recently in Africa and replaced Y chromosomes elsewhere in the world. Of course, this does not mean that other loci must show the same pattern, but the Y chromosome provides no evidence for genetic contributions from ancient *Homo* species, and few such signs have been detected using other loci.

Contemporary populations might therefore be largely the descendants of people who migrated out of Africa ~50 kya. Archaeological evidence indicates that there were at least two distinct migrations at this time. Populations using Middle Palaeolithic technology had probably reached Australia by ~50 kya[62], whereas distinct populations that used Upper Palaeolithic technology were present in Israel ~47 kya, Western Europe ~43 kya and sites in the Altai region of Russia ~42 kya[63,64]. This evidence therefore indicates that there might have been an early southern migration, perhaps following a coastal route around the northern edge of the Indian Ocean, before 50 kya, and a slightly later northern migration into Eurasia. To what extent does the pattern of Y-chromosome variation fit this model?

We might expect that the two migrations would carry distinct subsets of African Y chromosomes. There have been many subsequent migrations and demographic changes, but traces of overall differences might remain. The Y haplogroups that are found in southeastern Asia (considering the region broadly) and Australia are indeed distinct from those in much of the rest of Asia and Europe. Haplogroups C and O predominate, whereas D is common in some populations and M is frequent in New Guinea (FIG. 2). Haplogroup C has a wider distribution, which extends into Central and Northern Asia and the Americas, but D, M and O are rare outside southeastern Asia. The predominant haplogroups in northwestern Asia and Europe include I, J, N and R. Could the southern migration have carried C, D, M and O, and the northern carried I, J, N and R? Several of the haplogroups originated too recently to have been present at the time of these initial migrations, but their present distributions could reflect the earlier movements of their precursors. mtDNA, the only other locus for which comparable phylogenetic data are available, also shows a general distinction between southeastern Asia/Australia where mtDNA haplogroup M and its derivatives predominate, and northwestern Asia/Europe where N and its derivatives are more common.

A detailed analysis of Y-haplogroup C in Australia, however, cautions against interpreting all Y data as evidence of the earliest migrations[65]. Haplogroup C makes up approximately one-half of Australian Y chromosomes, but the limited microsatellite diversity of these, and their close similarity to C chromosomes from the Indian subcontinent, point to an entry into Australia during the HOLOCENE rather than >40 kya, which was perhaps associated with the introduction of the dingo and the changes in plant-processing techniques that took place at this later time. Nevertheless, the underlying worldwide patterns of Y-chromosome diversity can readily be interpreted within the archaeological framework, with four distinct sets of haplogroups in Africa (1), southeastern Asia/Australia (2), Central and Western Asia/Europe (3), and the regions that were settled later — the far north, the Americas and the Pacific (4).

Reshaping Y-chromosome diversity

Although these ancient traces can be discerned, many aspects of modern Y-chromosome diversity reflect subsequent events. There have been important changes in climate, which culminated in the last glacial maximum (~18–22 kya). Behaviourally-modern humans survived the ice age in many parts of the world, but only in refugia, which implies that there have been great contractions and expansions in range. The warm stable climate that developed during the Holocene allowed an agricultural way of life to be adopted after 10 kya, and was accompanied by large demographic expansions. Subsequent changes during the past few thousand years have been recorded by historians. What effects have these diverse events had on Y-chromosome variation? We do not have space for a comprehensive review of the field, but select a few studies that exemplify different principles. Many of these combine the clear SNP phylogeny with the resolving power of multiple microsatellites[66].

In Africa, the imprints of Palaeolithic events on Y-chromosome diversity are faint because of a recent and substantial expansion of haplogroup-E Y-chromosomes, which were probably carried by iron-working Bantu-speaking farmers from West Africa starting ~2.7 kya[67]. This is why haplogroups A and B are now rare, and most African Y chromosomes belong to haplogroup E. Also, there have been substantial movements into North Africa and there is evidence for at least one prehistoric migration back to sub-Saharan Africa, carrying haplogroup-R chromosomes to the Northern Cameroon region[60]. So, although recent events have been important, this rather simple view of African Y-chromosome diversity

HOLOCENE
The 'wholly recent' geological period that spans the past ~11,000 years and is characterized by an unusually warm and stable climate.

Box 5 | Applications in forensic and genealogical studies

DNA profiling using autosomal microsatellites is a sensitive and powerful method for linking crime scene to crime scene, and crime scene to suspect. In many countries, 'cold hits' to databases are helping to solve previously intractable crimes, and similar DNA-based technologies are used in paternity testing. Analysis of Y-chromosomal markers is not widely applied in these fields, but has specialized uses. For example, in rape cases, sperm cells are mixed with the epithelial cells of the victim and differential lysis usually allows sperm DNA to be isolated and profiled separately[101]. However, this method sometimes fails. Also, if a rapist is AZOOSPERMIC, then DNA must be extracted from the mixed epithelial cells of the victim and assailant. In such cases, Y-microsatellite analysis readily provides an assailant-specific profile[102]. In paternity testing of male children where the alleged father is unavailable, Y-chromosomal markers can be tested in male-line relatives, such as brothers[103]. Like any forensic DNA analysis, judging the significance of a match depends upon the population frequency of the profile. As has been discussed in the text, Y chromosomes are highly geographically structured and there is the further complication that all members of a patriline are expected to share Y haplotypes. The forensic community has been proactive in establishing large quality-assured and publicly accessible databases of Y-microsatellite haplotypes to address these issues[87,104] (see links to the Y-STR Haplotype Reference Database web sites in the online links box).

Although there is much overlap between the forensic and anthropological fields, there are also significant differences. Forensic databases are often substantially larger than anthropological ones, and precedence in legal casework means that adopting new markers is relatively slow. Also, forensic databases should be derived from a random set of individuals, whereas anthropologists go to considerable trouble to sample 'unrelated' 'indigenous' individuals. The implications of these different requirements, which can include the underestimation of common haplotype frequencies if anthropological data are used for forensic purposes, must be remembered in interdisciplinary analyses.

In many societies, surnames, like Y chromosomes, are patrilinearly inherited. This has led to considerable interest among amateur genealogists in whether branches of a family tree can be reliably connected using DNA evidence[105]. There is a burgeoning market in commercial DNA testing for genealogical purposes. However, the one published study on surnames and Y haplotypes[106] shows that the relationship between the two is not simple, and analyses of more surnames, together with further population data, are needed if reasonable interpretations are to be made.

undoubtedly reflects our lack of detailed information more than any real lack of complexity.

Populations outside Africa have received considerable attention, so we now have detailed information about the distribution of Y-chromosome diversity in Europe and parts of Asia. Factors that are important in influencing this diversity include population size and mobility. If populations are large and settled, change in haplogroup frequency owing to drift will be slow and local gene flow might lead to clinal patterns. These conditions could have been met in Holocene China and Western Asia/Europe, which were two important centres of agriculture. Detailed studies of Y-chromosome diversity in China are still awaited, but are available for Europe, where clinal patterns were found[68,69]. The most frequent haplogroup, R (xR1a), which account for 37% of the total sample, was present in all populations but was concentrated in the west, and was therefore interpreted as the haplogroup that was introduced by the first Palaeolithic inhabitants. Other clines in the frequency of haplogroups E3b, J, R1a and N3, pointed to sources to the south, southeast, east and northwest, respectively. The distribution of haplotype J fits archaeological data for the introduction of farming from the southeast and of N3 movement of Uralic speakers from the east, but correlates for E3b and R1a are less clear, which illustrates how genetic data can lead to new questions.

India has a large population, approximately one-sixth of the world total, which is highly structured by social factors such as the caste system, in which birth determines many aspects of life, including marriage partner. The large population is therefore subdivided into many ENDOGAMOUS subpopulations. There are hints that Y-chromosome diversity is highly structured along social lines[70], but also of haplotype sharing between groups[71,72], and more thorough studies are needed to understand the genetic consequences of such a social system.

Large regions of Central and Northern Asia are inhospitable and have probably never supported high population densities. Low male effective population size leads to strong genetic drift, so such regions show a different pattern of Y-chromosome diversity from more densely-populated areas. In a study of Siberia and adjacent regions, Karafet *et al.*[73] found that more than 96% of Siberian chromosomes fell into one of just four haplogroups, with many individual populations showing a single predominant lineage. Consequently, differences between populations, were well above the global average (with a Φ_{ST} of 0.41 compared with 0.36) and did not show a strong geographical pattern (for example, a cline) because different haplogroups had drifted to high frequency in nearby populations, and the effects of drift were not overcome by gene flow. Instead, and in contrast to Europe, genetic diversity was more structured according to language, which indicated that the Siberian populations, many of which are nomadic, might change their geographical location more often than their language. Similarly, in Central Asia, Zerjal *et al.*[74] found a marked distinction between populations in which the male effective population size was high and those in which it was low; in the latter, as in Siberia, a single haplogroup specific to the population had often drifted to high frequency. In the Kyrgyz and Kazak populations, the TMRCA of the predominant lineage corresponded to the time of the Mongol expansion under Genghis Khan in the thirteenth century, and might reflect the small number of males repopulating these areas. Although it is naive to view any modern population as similar to an ancient one, the patterns of Y-chromosome diversity seen in these regions might provide a better guide to those that prevailed in low-density prehistoric populations than the patterns now found in more populous areas.

The Americas were the last continents to be colonized. There is debate about the number of migrations, their timing and sources, but general agreement that the colonists came from or through northeast Asia, and had arrived by ~14.5 kya (~12.5 kya in uncalibrated radiocarbon years) according to evidence from Monte Verde in Chile[75]. Most Native American Y chromosomes belong to the single haplogroup Q (FIG. 2), and a high proportion fall into the single sublineage[76] within this, Q3 (REF. 77). As haplogroup Q is widespread in Central and Northern Asia (FIG. 2), but Q3 is confined to the Americas apart from

AZOOSPERMIA
The absence of sperm in the ejaculate.

ENDOGAMY
The practice of marrying within a social group.

Φ_{ST}
A measure of the subdivision between populations that takes into account the molecular distance between haplogroups/haplotypes, as well as their frequency.

LINKAGE DISEQUILIBRIUM
The non-random association between alleles in a population owing to their tendency to be co-inherited.

recent emigrants, the dates of the mutations defining these lineages should bracket the entry into the Americas. Dates of 17.7 ± 4.8 kya[40] and 7.6 ± 5 kya[78] have been estimated, which are consistent with the archaeological data and indicate that an early (>25 kya) entry of humans into the Americas is unlikely. The possibility has been raised that a small proportion of American Y-chromosome diversity originated in a second later migration, which introduced haplogroup C and some Q(xQ3) lineages that are found mainly in the north[78,79]. Multiple lineages do not necessarily imply multiple migrations, and post-1492 origins for some of these lineages from European admixture are possible[80], so the number of migrations remains unresolved. Most of the published studies were carried out before the polymorphisms that define haplogroup Q were available, so a re-examination of American chromosomes and the diversity and distribution of C and Q chromosomes in Asia, with increased phylogenetic resolution, will be valuable. Overall, Y-chromosome studies have provided important insights into the colonization of the Americas, and promise to provide more. They are, therefore, a model for how genetic analysis can contribute to controversial anthropological questions.

Perhaps the most unexpected finding in the field emerged from a study of Asian Y-chromosome diversity that, crucially, included 16 microsatellites among the markers used (BOX 4). As humans have increased in number, the patterns of diversity that were established largely by migration and drift during the Palaeolithic period have been 'frozen', and large-scale changes have become less frequent. Nevertheless, recent events can still occasionally have a significant influence on Y-chromosome diversity.

Comparisons of Y-chromosome data with mtDNA data have been particularly revealing about the sex-specific gene flow that accompanied the expansion of Europeans into the Americas and Oceania in the past 500 years. A typical pattern of strong introgression of European Y chromosomes with retention of indigenous mtDNA lineages is seen in Polynesia[81], Greenland[82,83] and South America[84–86], which reflects the sexual politics of colonial activity.

Conclusions

Development of the Y chromosome as an informative system for evolutionary studies in the 18 years since the description of the first DNA polymorphisms has been slow, but the field has now reached maturity. How could it develop in the future? Many Y-chromosome SNPs are available in databases, and although there are obstacles to their use, including the difficulty of distinguishing Y-chromosome polymorphisms from X–Y and Y–Y repeated sequence variants, and the ascertainment of these SNPs from a limited number of chromosomes, this resource is potentially valuable. Microsatellites can readily be identified from sequence information, and will probably soon be completely catalogued. This will provide the opportunity to use more such markers, and therefore to increase the precision of lineage discrimination to the point at which almost every Y chromosome, including those of fathers and sons, could be distinguished. It will also allow microsatellites to be used more selectively, so that, for example, those with simple mutational properties, or a particular level of variation, could be chosen. With such opportunities for increased data generation, the possibilities for confusion in the field also increase. The role of the YCC in maintaining and updating a standard phylogeny and nomenclature for binary markers and haplogroups remains crucial, but this would be usefully supplemented by a database that incorporated population-frequency data. The main users of Y-chromosome microsatellites are forensic scientists (BOX 5), who have chosen a standard 'core' set of seven markers to define a minimal haplotype (and nine for an extended haplotype)[87], and established databases of population haplotype frequencies for Europe, the United States and Asia. The number of microsatellites could now be increased: this has the advantage of improving discrimination but the disadvantage of requiring the expansion of existing databases. The development of powerful multiplexes that incorporate both existing and new microsatellites[88] provides one way to do this. The possibilities of incorporating binary markers, which might be amplified from highly degraded DNA, into forensic practice, and the prediction of binary haplogroup from microsatellite data (by anthropologists who want to use forensic data) deserve consideration.

Box 6 | A model for other regions of the genome

The phylogeographical information that is provided by the Y chromosome is powerful because it is based on detailed and readily interpreted haplotypes that are undisturbed by recombination. It has an obvious counterpart in mitochondrial DNA (mtDNA), but previously it seemed that the autosomes would provide little information because any segment long enough to contain many polymorphisms would recombine. However, there is now evidence for a haplotype-block structure to the genome: regions of strong LINKAGE DISEQUILIBRIUM (LD) on the autosomes. Blocks have been identified from the patterns of LD in populations[107] and some boundaries between blocks correspond to recombination hotspots[108]. Can each block be studied in the same way as the Y chromosome, using combinations of SNPs and microsatellites to define detailed haplotypes? This would address one of the main limitations of Y-chromosome studies: that they only provide information about a single locus, whereas many loci are needed to understand population processes. Some blocks identified in chromosome-wide analyses are more than 300 kb in length[109], and could potentially yield highly informative haplotypes.

This is an important area of investigation and there are grounds for optimism, but caution is necessary. Blocks could result predominantly from a combination of the chance distribution of past recombination events and drift; if so, they might differ substantially between populations and make it difficult to conduct a worldwide phylogeographic study. However, in at least one region of the genome — a 75 kb-long stretch in the *HLA* locus — blocks that are bounded by hotspots are shared between populations (United Kingdom Europeans, Zimbabweans and Saami) with different histories[110]. Although this region might be subject to selection, which would complicate its phylogeographic interpretation, searches for other regions that are contained in long hotspot-delineated blocks shared by all populations seem worthwhile. Even if only 1% of the genome were organized like this, 30 Mb of DNA would be available for study. A 100-kb block might contain several useful microsatellites and several hundred SNPs. Questions remain about the best way to determine or infer haplotypes from genotypes, and how to define blocks. The effects of errors in haplotypes and the best ways to interpret rare recombination events also need to be explored, but this promises to provide a rich source of data for future studies.

HAPLOTYPE BLOCK
The apparent haplotypic structure of the recombining portions of the genome, in which sets of consecutive co-inherited alleles are separated by short boundaries; there is debate about the origins of haplotype blocks and whether the boundaries correspond to recombination hotspots.

A more substantial change for anthropological purposes would be a move towards resequencing sections of all Y chromosomes in a study, rather than typing markers that were previously ascertained in a small (and often different) sample, with the attendant problems of ascertainment bias (BOX 1). Large-scale resequencing has been inhibited by the low level of Y-chromosome variation and the high cost, but as cost decreases, this will become more attractive. Studies of ancient DNA are, in principle, of enormous interest, but have been restricted to mtDNA because its high copy number makes it more likely to survive in an amplifiable form. Experiments on the *in vitro* repair of damaged nuclear DNA before PCR are promising[89], and might lead to reliable amplification of Y-chromosome sequences in the future.

The powerful phylogeographical information provided by the Y chromosome has a counterpart in mtDNA, and their comparison has often been particularly informative; however, it would be even better to examine a large number of independent loci. The identification of regions of low recombination on the autosomes (HAPLOTYPE BLOCKS) now indicates that this might be possible (BOX 6).

Sequencing of the chimpanzee genome is underway, and promises a cornucopia of information about the evolution of our own genome. Assembly of a chimp genome sequence using the human sequence as a framework will be straightforward for most chromosomes, but it might prove difficult for the Y chromosome because of its evolutionary lability. It is to be hoped that expenditure of effort on the Y chromosome will be comparable to that on other chromosomes, and that its reputation as a gene-poor junk-rich delinquent will not lead to a reluctance to include it wholeheartedly in the sequencing effort. It is notable (and sad) that the recently released draft genome sequence of the mouse[90] was derived from a female, and therefore entirely lacked data from a chromosome that, although undoubtedly a difficult 'customer', is nonetheless essential for the survival of the species.

1. Skaletsky, H. *et al.* The male-specific region of the human Y chromosome: a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003).
Analysis and evolutionary interpretation of the near-complete sequence of Y euchromatin, including thorough gene identification.
2. Bailey, J. A. *et al.* Recent segmental duplications in the human genome. *Science* **297**, 1003–1007 (2002).
3. Y Chromosome Consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* **12**, 339–348 (2002).
A collaborative effort by most of the groups in the field to produce a unified phylogeny and nomenclature, with a built-in updating mechanism, which is becoming widely adopted.
4. Hammer, M. F. A recent common ancestry for human Y chromosomes. *Nature* **378**, 376–378 (1995).
5. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97**, 7360–7365 (2000).
The largest study of Y-SNP variation free from ascertainment bias, which is based on DHPLC and proposes a common ancestor ~59 kya.
6. Murdock, G. P. *Ethnographic Atlas* (Univ. of Pittsburgh Press, Pittsburgh, 1967).
7. Burton, M. L., Moore, C. C., Whiting, J. W. M. & Romney, A. K. Regions based on social structure. *Curr. Anthropol.* **37**, 87–123 (1996).
8. Seielstad, M. T., Minch, E. & Cavalli-Sforza, L. L. Genetic evidence for a higher female migration rate in humans. *Nature Genet.* **20**, 278–280 (1998).
This paper proposed that patrilocality, the practice of a wife moving near to the birthplace of her husband after marriage rather than *vice versa*, could explain most of the excess geographical clustering of Y variants.
9. Kayser, M. *et al.* Independent histories of human Y chromosomes from Melanesia and Australia. *Am. J. Hum. Genet.* **68**, 173–190 (2001).
10. Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T. & Stoneking, M. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genet.* **29**, 20–21 (2001).
Empirical support for the hypothesis of Seielstad *et al.* in reference 8.
11. Nachman, M. W. & Crowell, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304 (2000).
12. Crow, J. F. The origins, patterns and implications of human spontaneous mutation. *Nature Rev. Genet.* **1**, 40–47 (2000).
13. Ebersberger, I., Metzler, D., Schwarz, C. & Pääbo, S. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* **70**, 1490–1497 (2002).
14. Blanco, P. *et al.* Divergent outcomes of intra-chromosomal recombination on the human Y chromosome: male infertility and recurrent polymorphism. *J. Med. Genet.* **37**, 752–758 (2000).
15. Kamp, C., Hirschmann, P., Voss, H., Huellen, K. & Vogt, P. H. Two long homologous retroviral sequence blocks in proximal Yq11 cause *AZFa* microdeletions as a result of intrachromosomal recombination events. *Hum. Mol. Genet.* **9**, 2563–2572 (2000).
16. Sun, C. *et al.* Deletion of azoospermia factor a (*AZFa*) region of human Y chromosome caused by recombination between HERV15 proviruses. *Hum. Mol. Genet.* **9**, 2291–2296 (2000).
17. Repping, S. *et al.* Recombination between palindromes P5 and P1 on the human Y chromosome causes massive deletions and spermatogenic failure. *Am. J. Hum. Genet.* **71**, 906–922 (2002).
18. Kuroda-Kawaguchi, T. *et al.* The *AZFc* region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature Genet.* **29**, 279–286 (2001).
19. Bosch, E. & Jobling, M. A. Duplications of the *AZFa* region of the human Y chromosome are mediated by homologous recombination between HERVs and are compatible with male fertility. *Hum. Mol. Genet.* **12**, 341–347 (2003).
20. Rozen, S. *et al.* Abundant gene conversion between arms of massive palindromes in human and ape Y chromosomes. *Nature* **423**, 873–876 (2003).
21. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. & de Knijff, P. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.* **6**, 799–803 (1997).
22. Kayser, M. *et al.* Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* **66**, 1580–1588 (2000).
Still the largest published study to measure the mutation rates at Y-chromosomal microsatellites using father–son pairs. This is more laborious than using deep-rooting pedigrees or sperm pools, but produces more reliable measurements.
23. Carvalho-Silva, D. R., Santos, F. R., Hutz, M. H., Salzano, F. M. & Pena, S. D. J. Divergent human Y-chromosome microsatellite evolution rates. *J. Mol. Evol.* **49**, 204–214 (1999).
24. Holtkemper, U., Rolf, B., Hohoff, C., Forster, P. & Brinkmann, B. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Hum. Mol. Genet.* **10**, 629–633 (2001).
25. Jeffreys, A. J., Murray, J. & Neumann, R. High-resolution mapping of crossovers in human sperm defines a minisatellite-associated recombination hotspot. *Mol. Cell* **2**, 267–273 (1998).
26. Jobling, M. A., Bouzekri, N. & Taylor, P. G. Hypervariable digital DNA codes for human paternal lineages: MVR-PCR at the Y-specific minisatellite, *MSY1* (*DYF155S1*). *Hum. Mol. Genet.* **7**, 643–653 (1998).
27. Jobling, M. A., Heyer, E., Dieltjes, P. & de Knijff, P. Y-chromosome-specific microsatellite mutation rates re-examined using a minisatellite, *MSY1*. *Hum. Mol. Genet.* **8**, 2117–2120 (1999).
28. Andreassen, R., Lundsted, J. & Olaisen, B. Mutation at minisatellite locus *DYF155S1*: allele length mutation rate is affected by age of progenitor. *Electrophoresis* **23**, 2377–2383 (2002).
29. Berta, P. *et al.* Genetic evidence equating *SRY* and the testis-determining factor. *Nature* **348**, 448–450 (1990).
30. Sun, C. *et al.* An azoospermic man with a *de novo* point mutation in the Y-chromosomal gene *USP9Y*. *Nature Genet.* **23**, 429–432 (1999).
31. Kuroki, Y. *et al.* Spermatogenic ability is different among males in different Y chromosome lineages. *J. Hum. Genet.* **44**, 289–292 (1999).
32. Carvalho, C. M. *et al.* Lack of association between Y chromosome haplogroups and male infertility in Japanese men. *Am. J. Med. Genet.* **116**, 152–158 (2003).
33. Previderé, C. *et al.* Y-chromosomal DNA haplotype differences in control and infertile Italian subpopulations. *Eur. J. Hum. Genet.* **7**, 733–736 (1999).
34. Jobling, M. A. *et al.* A selective difference between human Y-chromosomal DNA haplotypes. *Curr. Biol.* **8**, 1391–1394 (1998).
The first example of differential selection acting on Y haplogroups, which is supported by a plausible molecular mechanism.
35. Krausz, C. *et al.* Identification of a Y chromosome haplogroup associated with reduced sperm counts. *Hum. Mol. Genet.* **10**, 1873–1877 (2001).
36. Raitio, M. *et al.* Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. *Genome Res.* **11**, 471–482 (2001).
37. Paracchini, S., Arredi, B., Chalk, R. & Tyler-Smith, C. Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. *Nucleic Acids Res.* **30**, e27 (2002).
38. Zerjal, T. *et al.* The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72**, 717–721 (2003).
Selective expansion of a single Y haplotype, which is explained by a social rather than a biological mechanism.
39. Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Mol. Biol. Evol.* **16**, 1791–1798 (1999).
40. Hammer, M. F. & Zegura, S. L. The human Y chromosome haplogroup tree: nomenclature and phylogeny of its major divisions. *Annu. Rev. Anthropol.* **31**, 303–321 (2002).
41. Ingman, M., Kaessmann, H., Paabo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713 (2000).

42. Kaessmann, H., Heissig, F., von Haeseler, A. & Paabo, S. DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nature Genet.* **22**, 78–81 (1999).
43. Harris, E. E. & Hey, J. X chromosome evidence for ancient human histories. *Proc. Natl Acad. Sci. USA* **96**, 3320–3324 (1999).
44. Harding, R. M. *et al.* Archaic African *and* Asian lineages in the genetic ancestry of modern humans. *Am. J. Hum. Genet.* **60**, 772–789 (1997).
45. Tremblay, M. & Vezina, H. New estimates of intergenerational time intervals for the calculation of age and origins of mutations. *Am. J. Hum. Genet.* **66**, 651–658 (2000).
Genealogical research in Quebec, Canada, which shows that the male generation time in this population (~35 years) is longer than the female generation time (~29 years) and considerably longer than is commonly assumed (20–25 years).
46. Helgason, A., Hrafnkelsson, B., Gulcher, J. R., Ward, R. & Stefansson, K. A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am. J. Hum. Genet.* **72**, 1370–1389 (2003).
47. Brunet, M. *et al.* A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151 (2002).
48. Jobling, M. A. & Tyler-Smith, C. New uses for new haplotypes: the human Y chromosome, disease, and selection. *Trends Genet.* **16**, 356–362 (2000).
49. Shen, P. *et al.* Population genetic implications from sequence variation in four Y chromosome genes. *Proc. Natl Acad. Sci. USA* **97**, 7354–7359 (2000).
50. Hurles, M. E. *et al.* Native American Y chromosomes in Polynesia: the genetic impact of the Polynesian slave trade. *Am. J. Hum. Genet.* **72**, 1282–1287 (2003).
This paper shows the importance of taking historical information into account when interpreting genetic data.
51. Heyerdahl, T. *Kontiki: Across the Pacific by Raft* (Rand McNally, Chicago, 1950).
52. Swisher, C. C. *et al.* Age of the earliest known hominids in Java, Indonesia. *Science* **263**, 1118–1121 (1994).
53. Gabunia, L. *et al.* Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: taxonomy, geological setting, and age. *Science* **288**, 1019–1025 (2000).
54. Bermúdez de Castro, J. M. *et al.* A hominid from the lower Pleistocene of Atapuerca, Spain: possible ancestor to Neandertals and modern humans. *Science* **276**, 1392–1395 (1997).
55. White, T. D. *et al.* Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**, 742–747 (2003).
56. Valladas, H. *et al.* Thermoluminescence dating of Mousterian 'proto-Cro-Magnon' remains from Israel and the origin of modern man. *Nature* **331**, 614–616 (1988).
57. Valladas, H. *et al.* Thermoluminescence dates for the Neanderthal burial site at Kebara in Israel. *Nature* **330**, 159–160 (1987).
58. Mellars, P. Archaeology and the origins of modern humans: European and African perspectives. *Proc. Brit. Acad.* **106**, 31–47 (2002).
59. Hammer, M. F. *et al.* Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Mol. Biol. Evol.* **15**, 427–441 (1998).
A sophisticated model-based approach to testing hypotheses to explain patterns of global Y diversity.
60. Cruciani, F. *et al.* A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. *Am. J. Hum. Genet.* **70**, 1197–1214 (2002).
61. Underhill, P. A. *et al.* Y chromosome sequence variation and the history of human populations. *Nature Genet.* **26**, 358–361 (2000).
This study introduced 160 new binary Y markers and revealed much of the shape of the Y phylogeny and the geographical distribution of lineages.
62. Bowler, J. M. *et al.* New ages for human occupation and climatic change at Lake Mungo, Australia. *Nature* **421**, 837–840 (2003).
63. Goebel, T. Pleistocene human colonization of Siberia and peopling of the Americas: an ecological approach. *Evol. Anthropol.* **8**, 208–227 (1999).
64. Mellars, P. in *The Speciation of Modern Homo sapiens* (ed. Crow, T. J.) 31–47 (Oxford Univ. Press, Oxford, UK, 2002).
65. Redd, A. J. *et al.* Gene flow from the Indian subcontinent to Australia: evidence from the Y chromosome. *Curr. Biol.* **12**, 673–677 (2002).
66. Zerjal, T. *et al.* Genetic relationships of Asians and northern Europeans, revealed by Y-chromosomal DNA analysis. *Am. J. Hum. Genet.* **60**, 1174–1183 (1997).
67. Phillipson, D. W. *African Archaeology* 2nd edn (Cambridge Univ. Press, Cambridge, 1993).
68. Rosser, Z. H. *et al.* Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *Am. J. Hum. Genet.* **67**, 1526–1543 (2000).
69. Semino, O. *et al.* The genetic legacy of Paleolithic *Homo sapiens sapiens* in extant Europeans: a Y chromosome perspective. *Science* **290**, 1155–1159 (2000).
70. Bhattacharyya, N. P. *et al.* Negligible male gene flow across ethnic boundaries in India, revealed by analysis of Y-chromosomal DNA polymorphisms. *Genome Res.* **9**, 711–719 (1999).
71. Ramana, G. V. *et al.* Y-chromosome SNP haplotypes suggest evidence of gene flow among caste, tribe, and the migrant Siddi populations of Andhra Pradesh, South India. *Eur. J. Hum. Genet.* **9**, 695–700 (2001).
72. Kivisild, T. *et al.* The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am. J. Hum. Genet.* **72**, 313–332 (2003).
73. Karafet, T. M. *et al.* High levels of Y-chromosome differentiation among native Siberian populations and the genetic signature of a boreal hunter–gatherer way of life. *Hum. Biol.* **74**, 761–789 (2002).
A good example of what can be done with the present set of markers and analytical tools.
74. Zerjal, T., Wells, R. S., Yuldasheva, N., Ruzibakiev, R. & Tyler-Smith, C. A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *Am. J. Hum. Genet.* **71**, 466–482 (2002).
75. Meltzer, D. A. Monte Verde and the Pleistocene peopling of the Americas. *Science* **276**, 754–755 (1997).
76. Pena, S. D. *et al.* A major founder Y-chromosome haplotype in Amerindians. *Nature Genet.* **11**, 15–16 (1995).
77. Underhill, P. A., Jin, L., Zemans, R., Oefner, P. J. & Cavalli-Sforza, L. L. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl Acad. Sci. USA* **93**, 196–200 (1996).
78. Karafet, T. M. *et al.* Ancestral Asian source(s) of New World Y-chromosome founder haplotypes. *Am. J. Hum. Genet.* **64**, 817–831 (1999).
79. Lell, J. T. *et al.* The dual origin and Siberian affinities of Native American Y chromosomes. *Am. J. Hum. Genet.* **70**, 192–206 (2002).
80. Tarazona-Santos, E. & Santos, F. R. The peopling of the Americas: a second major migration? *Am. J. Hum. Genet.* **70**, 1377–1380 (2002).
81. Hurles, M. E. *et al.* European Y-chromosomal lineages in Polynesia: a contrast to the population structure revealed by mitochondrial DNA. *Am. J. Hum. Genet.* **63**, 1793–1806 (1998).
82. Saillard, J., Forster, P., Lynnerup, N., Bandelt, H.-J. & Nørby, S. mtDNA variation among Greenland Eskimos: the edge of the Beringian expansion. *Am. J. Hum. Genet.* **67**, 718–726 (2000).
83. Bosch, E. *et al.* High level of male-biased Scandinavian admixture in Greenlandic Inuit shown by Y-chromosomal analysis. *Hum. Genet.* **112**, 353–363 (2003).
84. Alves-Silva, J. *et al.* The ancestry of Brazilian mtDNA lineages. *Am. J. Hum. Genet.* **67**, 444–461 (2000).
85. Carvalho-Silva, D. R., Santos, F. R., Rocha, J. & Pena, S. D. J. The phylogeography of Brazilian Y-chromosome lineages. *Am. J. Hum. Genet.* **68**, 281–286 (2001).
86. Carvajal-Carmona, L. G. *et al.* Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. *Am. J. Hum. Genet.* **67**, 1287–1295 (2000).
87. Roewer, L. *et al.* Online reference database of Y-chromosomal short tandem repeat (STR) haplotypes. *Forens. Sci. Int.* **118**, 103–111 (2001).
88. Butler, J. M. *et al.* A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. *Forensic Sci. Int.* **129**, 10–24 (2002).
89. Di Bernardo, G. *et al.* Enzymatic repair of selected cross-linked homoduplex molecules enhances nuclear gene rescue from Pompeii and Herculaneum remains. *Nucl. Acids Res.* **30**, e16 (2002).
90. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
91. Su, B. *et al.* Y-chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age. *Am. J. Hum. Genet.* **65**, 1718–1724 (1999).
92. Karafet, T. *et al.* Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *Am. J. Hum. Genet.* **69**, 615–628 (2001).
93. Eyre-Walker, A. & Awadalla, P. Does human mtDNA recombine? *J. Mol. Evol.* **53**, 430–435 (2001).
94. Walzer, S. & Gerald, P. S. A chromosome survey of 13,751 male newborns. in *Population Cytogenetics* (eds Hook, E. B. & Porter, I. H.) 45–61 (Academic Press, New York, 1977).
95. Chevret, E. *et al.* Meiotic behaviour of sex chromosomes investigated by three-colour FISH on 35 142 sperm nuclei from two 47,XYY males. *Hum. Genet.* **99**, 407–412 (1997).
96. Solari, A. J. & Rey-Valzacchi, G. The prevalence of a YY synaptonemal complex over XY synapsis in an XYY man with exclusive XYY spermatocytes. *Chrom. Res.* **5**, 467–474 (1997).
97. Pecon-Slattery, J., Sanner-Wachter, L. & O'Brien, S. J. Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). *Proc. Natl Acad. Sci. USA* **97**, 5307–5712 (2000).
98. Cooke, H. J. & Noel, B. Confirmation of Y/autosome translocations using recombinant DNA. *Hum. Genet.* **50**, 39–44 (1979).
99. Andersson, M. *et al.* Y;autosome translocations and mosaicism in the aetiology of 45,X maleness: assignment of fertility factor to distal Yq11. *Hum. Genet.* **79**, 2–7 (1988).
100. Juvaini, A.-M. *The History of the World-Conqueror* (UNESCO publishing, 1260) (translated in 1997).
101. Gill, P., Jeffreys, A. J. & Werrett, D. J. Forensic application of DNA 'fingerprints'. *Nature* **318**, 577–579 (1985).
102. Sibille, I. *et al.* Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa. *Forens. Sci. Int.* **125**, 212–216 (2002).
103. Rolf, B., Keil, W., Brinkmann, B., Roewer, L. & Fimmers, R. Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations. *Int. J. Legal Med.* **115**, 12–15 (2001).
104. Kayser, M. *et al.* Online Y-chromosomal short tandem repeat haplotype reference database (YHRD) for US populations. *J. Forens. Sci.* **47**, 513–519 (2002).
105. Jobling, M. A. In the name of the father: surnames and genetics. *Trends Genet.* **17**, 353–357 (2001).
106. Sykes, B. & Irven, C. Surnames and the Y chromosome. *Am. J. Hum. Genet.* **66**, 1417–1419 (2000).
107. Daly, M. J., Rioux, J. D., Schaffner, S. F., Hudson, T. J. & Lander, E. S. High-resolution haplotype structure in the human genome. *Nature Genet.* **29**, 229–232 (2001).
108. Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet.* **29**, 217–222 (2001).
109. Phillips, M. S. *et al.* Chromosome-wide distribution of haplotype blocks and the role of recombination hot spots. *Nature Genet.* **33**, 382–387 (2003).
110. Kauppi, L., Sajantila, A. & Jeffreys, A. J. Recombination hotspots rather than population history dominate linkage disequilibrium in the MHC class II region. *Hum. Mol. Genet.* **12**, 33–40 (2003).
111. Santos, F. R., Pandya, A. & Tyler-Smith, C. Reliability of DNA-based sex tests. *Nature Genet*. **18**, 103 (1998).
112. Semino, O., Santachiara-Benerecetti, A. S., Falaschi, F., Cavalli-Sforza, L. L. & Underhill, P. A. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. *Am. J. Hum. Genet.* **70**, 265–268 (2002).
113. Wells, R. S. *et al.* The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc. Natl Acad. Sci. USA* **98**, 10244–10249 (2001).
114. Qian, Y. *et al.* Multiple origins of Tibetan Y chromosomes. *Hum. Genet.* **106**, 453–454 (2000).
115. Kayser, M. *et al.* Reduced Y-chromosome, but not mitochondrial DNA, diversity in human populations from West New Guinea. *Am. J. Hum. Genet.* **72**, 281–302 (2003).
116. Capelli, C. *et al.* A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. *Am. J. Hum. Genet.* **68**, 432–443 (2001).
117. Underhill, P. A. *et al.* Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum. Mutat.* **17**, 271–280 (2001).
118. Tarazona-Santos, E. *et al.* Genetic differentiation in South Amerindians is related to environmental and cultural diversity: evidence from the Y chromosome. *Am. J. Hum. Genet.* **68**, 1485–1496 (2001).
119. Thangaraj, K. *et al.* Genetic affinities of the Andaman islanders, a vanishing human population. *Curr. Biol.* **13**, 86–93 (2003).
120. Paracchini, S. *et al.* Y-chromosomal DNA haplotypes in infertile European males carrying Y-microdeletions. *J. Endocrinol. Invest.* **23**, 671–676 (2000).

121. Quintana-Murci, L. *et al.* The relationship between Y chromosome DNA haplotypes and Y chromosome deletions leading to male infertility. *Hum. Genet.* **108**, 55–58 (2001).
122. Paracchini, S. *et al.* Relationship between Y-chromosomal DNA haplotype and sperm count in Italy. *J. Endocrinol. Invest.* **25**, 993–995 (2002).
123. Kittles, R. A. *et al.* Cladistic association analysis of Y chromosome effects on alcohol dependence and related personality traits. *Proc. Natl Acad. Sci. USA* **96**, 4204–4209 (1999).
124. Ellis, J. A., Stebbing, M. & Harrap, S. B. Significant population variation in adult male height associated with the Y chromosome and the aromatase gene. *J. Clin. Endocrinol. Metab.* **86**, 4147–4150 (2001).
125. Charchar, F. J. *et al.* The Y chromosome effect on blood pressure in two European populations. *Hypertension* **39**, 353–356 (2002).
126. Ellis, J. A., Stebbing, M. & Harrap, S. B. Association of the human Y chromosome with high blood pressure in the general population. *Hypertension* **36**, 731–733 (2000).
127. Shoji, M. *et al.* Lack of association between Y chromosome Alu insertion polymorphism and hypertension. *Hypertens. Res.* **25**, 1–3 (2002).
128. Ewis, A. A. *et al.* Linkage between prostate cancer incidence and different alleles of the human Y-linked tetranucleotide polymorphism *DYS19*. *J. Med. Invest.* **49**, 56–60 (2002).
129. Quintana-Murci, L. *et al.* Y chromosome haplotypes and testicular cancer in the English population. *J. Med. Genet.* **40**, e20 (2003).
130. Passarino, G. *et al.* Y chromosome binary markers to study the high prevalence of males in Sardinian centenarians and the genetic structure of the Sardinian population. *Hum. Hered.* **52**, 136–139 (2001).
131. Jamain, S. *et al.* Y chromosome haplogroups in autistic subjects. *Mol. Psychiatry* **7**, 217–219 (2002).

Acknowledgements

We thank T. Karafet for unpublished information on American populations, the Y Chromosome Consortium for consultation and for allowing us to include figure 3, and M. Hurles, F. Santos and three anonymous reviewers for helpful comments on the manuscript. M.A.J. is supported by a Wellcome Trust Senior Fellowship in Basic Biomedical Science. We apologize to those collegues whose work we could not cite because of limited space.

Online links

DATABASES

The following terms in this article are linked online to:

LocusLink: http://www.ncbi.nlm.nih.gov/LocusLink

PRKX | *PRKY* | *USP9Y*

FURTHER INFORMATION

BATWING program:
http://maths.abdn.ac.uk/~ijw/downloads/download.htm

Ensembl: http://www.ensembl.org

Phylogenetic Network Analysis Shareware Software:
http://www.fluxus-engineering.com/sharenet.htm

Y Chromosome Consortium:
http://ycc.biosci.arizona.edu

Y-STR Haplotype Reference Database Asia:
http://ystr.org/asia

Y-STR Haplotype Reference Database Europe:
http://ystr.org/europe

Y-STR Haplotype Reference Database for US Populations:
http://ystr.org/usa

Access to this interactive links box is free online.