A singular chromosome

Matthew E Hurles & Mark A Jobling

The human Y chromosome is the first constitutively haploid metazoan chromosome to be sequenced. It has a unique genomic landscape with a complex evolutionary history that has endowed it with few genes but many nearly identical dispersed repeats that underlie the structural fluidity of this unusual chromosome.

In a way, it's Randy's fault. One-time cover-star of *Nature*, this trans-sexual transgenic mouse, carrying a mere 14 kb of Y chromosome but nevertheless evidently a boy[1], entrenched the idea that the mammalian Y chromosome was a one-trick, one-gene show. The *SRY* gene confers testes, and the other chromosomes do the rest. The recent publication in *Nature* of the sequence of the male-specific portion of the human Y chromosome (MSY; ref. 2) presents a different picture, however, of an extraordinary part of the genome shaped by a strange history of haploidy.

The MSY constitutes about 90% of the euchromatic length of the human Y chromosome, with the remainder residing in pseudoautosomal regions that recombine with counterparts on the tips of the X chromosome. The MSY passes only through the male germ line, and its haploid nature precludes allelic recombination during meiosis. These factors should have had a profound influence on the chromosomal composition, and the MSY sequence shows just how markedly this chromosome differs from genomic norms[2].

Evolution of a gene-poor chromosome

The mammalian X and Y chromosomes derive from a pair of autosomes, and the acquisition of a sex-determining function by one of these initiated their divergence[3]. The sex-determining function resided in the nascent MSY, where recombination between proto-sex chromosomes was repressed. The MSY has progressively expanded at the expense of the pseudoautosomal regions[4]; genes now in the MSY acquired male-specificity at different times, owing to sequential chromosomal rearrangements that imposed male-specificity on previously pseudoautosomal sequences by repressing XY recombination. The clear relationship between the dates at which ancestral XY genes stopped recombining and their X-chromosomal positions suggests that Y-chromosomal rearrangements were predominant in disrupting pseudoautosomal recombination. This idea is supported by the observation of marked karyotypic variation among Y chromosomes in primates[5].

What genes are found on the only chromosome that half our species lives quite happily without? *In silico* and RT–PCR analyses of the 23 Mb of sequence identified 156 transcription units, of which only 78 seem to encode proteins. These 78 belong to only 27 gene families, with very high intrafamily similarity, so that distinct protein-encoding genes are few and far between (~10% of the genome-average density). They fall into two main classes: 16 ubiquitously expressed single-copy genes found in portions of the chromosome that derive from the proto-sex chromosomes (classified as 'X-degenerate' by the authors) and 9 testis-specific multi-copy genes with diverse chromosomal origins found exclusively in segmentally duplicated portions of the chromosome (confusingly named 'ampliconic'). Population geneticists[6] have suggested that genes with male-specific functions should accumulate in male-specific portions of the genome. This is indicated by the presence on MSY of testis-specific genes (several of which have known spermatogenic functions) and by the diversity of their origins. But the inevitable degeneration of non-recombining sequences means that these acquisitions are outnumbered by the loss of most Y-chromosomal copies of the genes present on the proto-sex chromosomes. Furthermore, the recent human-specific transposition of 3.4 Mb of gene-poor sequence from the X to the Y (termed 'X-transposed') has added only two new genes.

Multiple segmental duplications

One surprising finding of the human genome project was the high proportion (~5%) of sequence present in large (>10 kb) duplicated segments[7]. In this context, three features of the pattern of segmental duplication in the MSY stand out. First, the sheer proportion of the MSY present in segmental duplications (30–45% depending on how it is defined) contrasts sharply with the genome average. Navigating across the larger segmental duplications necessitated a painstaking combination of mapping and sequencing, which may act as a model for analyzing the segmental duplication–rich pericentromeric regions of other chromosomes that are largely uncharacterized[8]. Second, most MSY segmental duplications share >99.5% sequence similarity, whereas in the rest of the genome, sequence similarity between segmental duplications is more broadly distributed between 90% and 100% (ref. 9). Finally, almost all of this nearly identical class of segmental duplication is present in tightly spaced palindromes. The very high sequence similarity between these palindromes suggests either that they are very recent in origin or that concerted evolution[10] is acting to maintain similarity between intra-specific repeats. In an accompanying paper in *Nature*[11], the existence of chimpanzee homologs shows that these palindromes are substantially older than their minimal divergence would have us believe and that the recent-origin hypothesis is wrong. Comparative sequencing data suggest that abundant gene conversion (the non-reciprocal transfer of sequence from one homologous sequence to another) has driven the concerted evolution of these palindrome arms[11].

The authors put forward a contentious hypothesis that these palindromic sequences have a functional role in promoting gene conversion that protects essential genes against the degeneration that is the inevitable consequence of haploidy. This idea that safety in numbers is conferred by duplication, followed by gene conversion biased towards preserving the wild-type sequence[11], requires careful modeling and scrutiny before it can become widely accepted.

There is, however, a negative aspect to this high density of segmental duplication. Gene conversion is but one outcome of homologous recombination between duplicated sequences, the other being crossover resulting in deletion, duplication

M.E.H. is in the McDonald Institute for Archaeological Research, University of Cambridge, Downing Street, Cambridge CB2 3ER, UK. M.A.J. is in the Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK. e-mail: meh32@cam.ac.uk or maj4@le.ac.uk

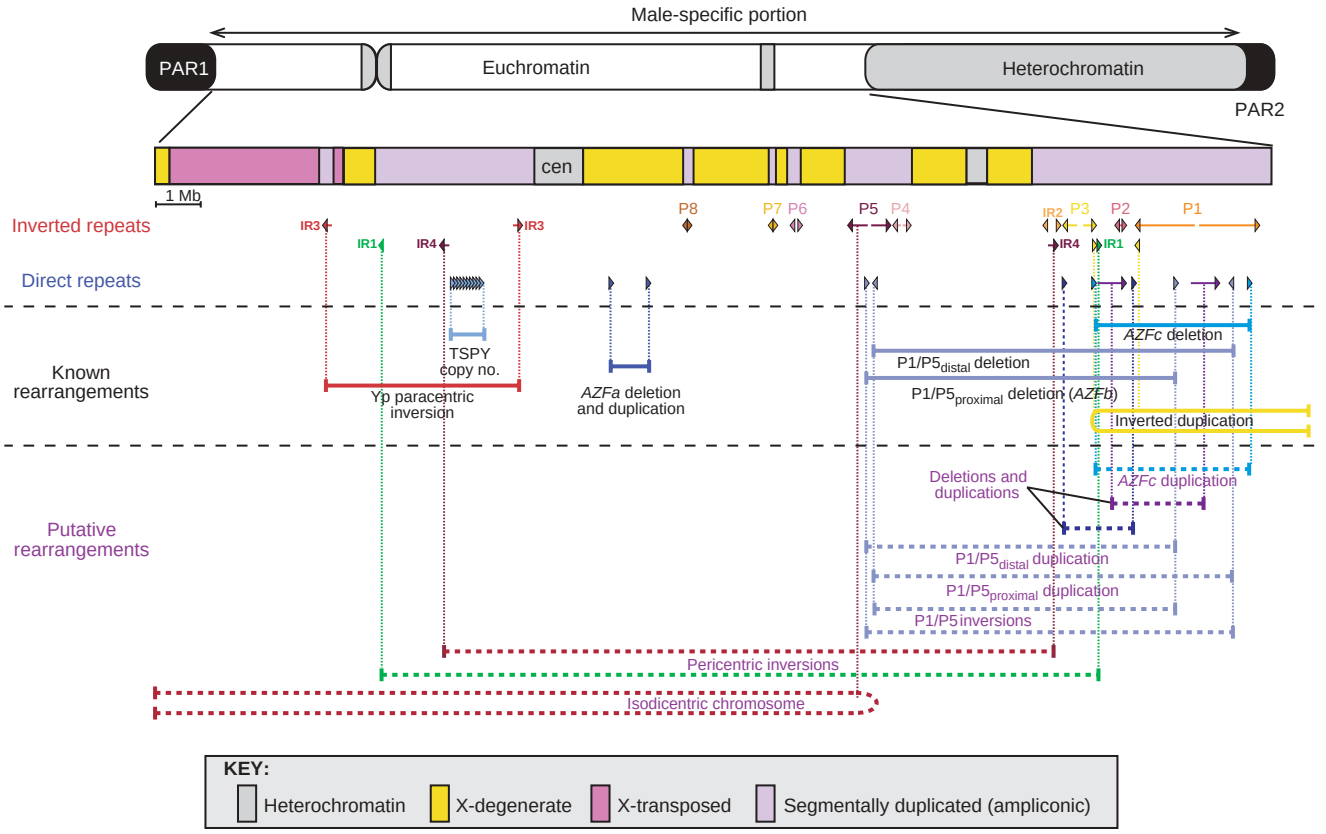

Figure 1 Sequence organization of the human Y chromosome. The mosaic of three main sequence classes is shown together with the principal inverted and direct repeats on the MSY. Eight palindromes (P1–P8) and four dispersed inverted repeats (IR1–IR4) are labeled. Beneath these are shown the known rearrangements, and a representative subset of putative rearrangements, that are associated with these repeats.

or inversion of intervening sequence[12]. The MSY harbors a variety of dispersed segmental duplications that sponsor a wide range of pathogenic rearrangements and structural polymorphisms[13–15] (**Fig. 1**). The high frequency of Y-chromosomal rearrangement may result in part from the lack of a meiotic pairing partner that leaves the chromosome free to fold back on itself. This hypothesis is supported by the observation that over 90% of *de novo* hemophilia inversions on the X chromosome occur during male meioses[16].

The remarkable properties of the human Y chromosome are now becoming clear, but the MSY sequence raises as many questions as it answers. For example, does the predominance of palindromic segmental duplications result from biases in duplication processes or selection against direct repeats that are more likely to predispose to pathogenic rearrangements? And what is the function of the many apparently untranslated, though spliced, transcription units? A large-scale comparison with the chimpanzee Y chromosome should prove insightful. The sequences of other haploid chromosomes, particularly those from species (like birds) in which the heterogametic sex is the female, will throw further light on the strange consequences of the haploid life. It's unfortunate that the Mouse Genome Sequencing Consortium chose a female, rather than a male, for sequencing and so omitted the Y chromosome from the genome sequence of the standard mammalian animal model. Maybe Randy is at it again.

1. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P. & Lovell-Badge, R. *Nature* **351**, 117–121 (1991).
2. Skaletsky, H. *et al. Nature* **423**, 825–837 (2003).
3. Ohno, S. *Sex chromosomes and sex-linked genes* (Springer, Berlin, 1967).
4. Lahn, B.T. & Page, D.C. *Science* **286**, 964–967 (1999).
5. Glaser, B. *et al. Mamm. Genome* **9**, 226–231 (1998).
6. Fisher, R.A. *Biol. Rev.* **6**, 345–368 (1931).
7. Bailey, J.A. *et al. Science* **297**, 1003–1007 (2002).
8. Guy, J. *et al. Genome Res.* **13**, 159–172 (2003).
9. Samonte, R.V. & Eichler, E.E. *Nat. Rev. Genet.* **3**, 65–72 (2002).
10. Zimmer, E.A., Martin, S.L., Beverley, S.M., Kan, Y.W. & Wilson, A.C. *Proc. Natl. Acad. Sci. USA* **77**, 2158–2162 (1980).
11. Rozen, S. *et al. Nature* **423**, 873–876 (2003).
12. Stankiewicz, P. & Lupski, J.R. *Trends Genet.* **18**, 74–82 (2002).
13. Blanco, P. *et al. J. Med. Genet.* **37**, 752–758 (2000).
14. Bosch, E. & Jobling, M.A. *Hum. Mol. Genet.* **12**, 341–347 (2003).
15. Repping, S. *et al. Mol. Hum. Reprod.* **9**, 183–188 (2003).
16. Rossiter, J.P. *et al. Hum. Mol. Genet.* **3**, 1035–1039 (1994).