Cell, Vol. 113, 825–828, June 27, 2003, Copyright ©2003 by Cell Press

The Human Y Chromosome: Rumors of Its Death Have Been Greatly Exaggerated

Minireview

R. Scott Hawley[1,2,*]
[1]Stowers Institute for Medical Research
Kansas City, Missouri 64110
[2]Department of Molecular Biosciences
University of Kansas
Lawrence, Kansas 66045

Recent papers by David Page and his collaborators present an intriguing new face to the human Y chromosome, including eight massive palindromic arrays, most of which contain multi-copy pairs of testis-specific genes. Sequence pairs within the arms of these palindromic sequences retain a high degree of similarity, suggesting that intrachromosomal gene conversion is frequent, perhaps providing a means of maintaining the integrity of individual members of the array.

The Y chromosome is oft perceived as a genetic "wrecking yard," housing the rusting chassis of genes that can no longer be driven. Presumably once a proud and fully functional homolog to the X chromosome, it is now thought by some to ferry only the gene that determines maleness (SRY) and a small number of other protein-encoding genes whose long-term evolutionary doom seems all but certain. Indeed, at least some investigators describe the human Y chromosome as a "wimp" among chromosomes and predict the demise of the entire chromosome within 5–10 million years (Aitken and Marshall Graves, 2002; Marshall Graves, 2000). The reason for this rapid decay of things Y-chromosomal is thought to be quite simple: once the Y chromosome became sex-determining, its presence was limited to the heterogametic sex (in our case, males). Because the Y chromosome was never found in the absence of an X chromosome, there was presumably little selection against the mutational inactivation of those genes on the Y chromosome that were also present on the X chromosome. Thus, over evolutionary time, the Y chromosome gradually lost most of its functional genes by the accumulation of deleterious mutations, resulting in that little dab of male-determining chromatin that we have today.

This rather desolate picture of the Y chromosome is consistent with its gross anatomy (Lahn et al., 2001). It is a largely heterochromatic and gene-poor chromosome that recombines with its homolog only at its ends, the so-called pseudoautosomal regions (see Figure 1). However, the pseudoautosomal regions comprise only 5% of the Y chromosome and contain only a dozen or so genes (all of which are also found on the X chromosome). The majority (63 Mb) of the Y chromosome (known as the male specific Y or MSY region) does not recombine with the X chromosome, and approximately two-thirds (41Mb) of this region is comprised of three blocks of highly reiterated satellite sequence. Even the 23 Mb of the MSY that is euchromatic appeared to be a primarily a wasteland of degenerating X chromosomal genes. Certainly, studies of infertile males who carry Y chromosomal deletions had revealed the existence of a few genes in this interval that appeared to be present only on the Y and which were required for male-fertility (Reijo et al., 2000), but such genes were few in number and even their evolutionary future seemed uncertain.

Novel Y Chromosome Elements that Appear Quite Vigorous Indeed

Fortunately for those of us who value this chromosome, David Page and his collaborators have identified a set of new structural elements within the euchromatic MSY that would appear to have a bit of evolutionary "staying power" (Skaletsky et al., 2003). This set of seven regions, referred to as the ampliconic regions, has a combined length of 10.2 Mb. It contains nine families of Y-specific protein-coding genes whose copy number ranges from 2 to 35 (totaling 60 genes altogether). All nine of these protein-encoding gene families are specifically expressed in the testis. There are also 75 transcription units that may or may not encode proteins, of which 65 are members of gene families whose members are entirely found on MSY. In total, these 135 ampliconic transcription units represent the vast majority of the 156 transcription units found within the entire MSY region. But the most impressive feature of these ampliconic regions is the sequence identity within the long repeated elements that comprise them. Most of the ampliconic regions are made up of long sequences that display 99.9% identity to other sequences in the ampliconic region. Indeed, the degree of sequence identity between various repeated sequences is so high that the sequence analysis had to be performed using Y chromosomal DNA from a single male, since the degree of polymorphism between males exceeded the degree of polymorphism between repeats.

Thus, far from being a genetic wasteland, the ampliconic regions of the Y appear to represent regions of multi-copy testis-specific genes whose sequences are highly homogenous. This finding leads to two questions: namely, how are these regions organized, and what mechanisms account for the high degree of sequence conservation? Much of the ampliconic region on the long arm of the Y is made up of eight giant palindromes (their arms range in size from 9 kb to 1.45 Mb) that in total make up some 25% of the euchromatic MSY. The largest of these palindromes (denoted P1) has a total length of 2.9 Mb and possesses two secondary palindromes within its arms. These palindromic regions contain members of eight of the nine families of protein-encoding genes, and six of these families are located exclusively on the palindromes, including the much studied DAZ gene family whose products appears to be required at multiple points during male germ cell development (Reijo et al., 2000). At least six of these palindromes predate the divergence of humans and chimpanzees (approximately 5 million years ago). The ampliconic regions also contain five sets of widely spaced inverted repeats, ranging in length from 62 to 298 kb as well as a variety of long tandem repeats, the most notable of which are 622 and 700 bp in length.

*Correspondence: rsh@stowers-institute.org

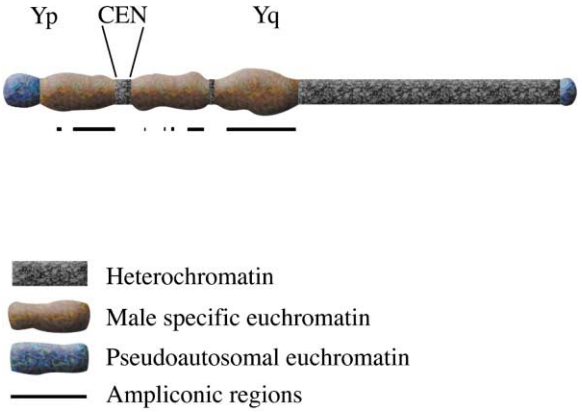

Figure 1. A Schematic Drawing of the Human Y Chromosomes Denoting the Regions of Eu- and Heterochromatin, as Well as the Pseudoautosomal Regions

The locations of the ampliconic regions are denoted by black horizontal bars underneath the chromosome. The symbols Yp and Yq denote the short and long arms of the Y chromosome, respectively.

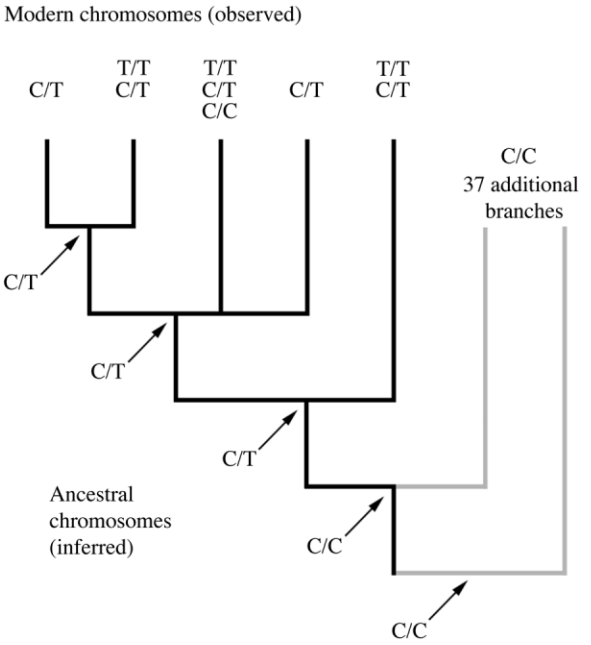

Figure 2. Evidence for Gene Conversion between Homologous Sequences on Opposite Arms of a Y Chromosome Palindrome

A genealogical tree of a Y chromosome lineage in which individual chromosomes denoting the sequence differences between two homologous sites on opposite arms of a palindrome. As described in the text, males are defined as either C/C, C/T, or T/T depending on the nucleotides present at the polymorphic sites in two copies of the CDY gene located at the same position on opposite arms of the palindrome. Thus, individuals with a C at the same position on opposite arms are denoted as C/C, those with a C on one arm and a T on other are denoted as C/T, and so on. Thanks are due to Ed vanVeen for creating both of these figures.

The 700 kb repeat includes 35 copies of the testis-specific TSPY gene.

Sequence Homogeneity with the Ampliconic Regions Is Maintained by Intrachromosomal Gene Conversion Events

Several lines of evidence suggest that the high degree of sequence identity between the arms of the palindromes and within the tandem repeats reflects the result of frequent gene conversion events. Most notably, studies of two copies of the CDY gene that are located at the same positions on opposite arms of P1 provide clear evidence of homogenization of a sequence polymorphism within one Y chromosome lineage (Rozen et al., 2003). Page and his collaborators sequenced these two CDY genes in 171 men, chosen to represent a tree of Y chromosomal genealogy with 42 branches. In doing so, they identified some Y chromosomes with a C at the same site on both arms of the palindrome, other Y chromosomes with a T at the same sites on opposite arms of the palindrome, and still others with a C at one site and a T at the other (see Figure 2). However, the C/T and T/T –bearing Y chromosomes were confined to a young cluster of five closely-related branches. The most parsimonious explanation for these observations is that the Y chromosome from which these five branches arose carried a C/T polymorphism, arising as a result of a C → T transition mutation. Both T/T and C/T chromosomes were observed within the young cluster of the five branches, and C/C chromosomes were observed in one of the five branches. To quote the authors, "thus, during recent human history, gene conversion in C/T chromosomes has used either the C copy or the T copy as a template." Similar observations were made at two other duplicated sites that displayed sequence variation. Moreover, an analysis of the ampliconic sequences of the chimpanzee and the human lends further support to the idea of gene conversion events occurring in these regions. Within each species, comparison of intrachromosomal repeats revealed very little sequence variation (usually less than .03%). In contrast, the degree of inter-species variation between the same regions of the Y chromosomes of humans and chimpanzees was significantly higher (1.4 to 2.3%, depending on the regions compared), suggesting that homogenization within ampliconic regions has continued to occur within the history of each species since the time of their divergence (Rozen et al., 2003).

So how frequent are gene conversion events in the ampliconic regions? The near uniformity of arm-to arm sequence in the palindromes of both humans and chimpanzees suggests that gene conversion must be frequent enough to erase new mutations almost as quickly as they occur. Using a mutation rate of 1×10^{-9} substitutions per nucleotide per year, the authors calculate that on average "600 duplicated nucleotides have undergone arm-to-arm gene conversion for every son born in recent human population." It is not clear whether or not these events are meiotic or mitotic, but it is obvious that these regions of the Y chromosome are not recombinationally inert, as had been previously thought. They may not recombine with a homolog, but intra-chromosomal gene conversion is certainly frequent!

All of this raises the question of the role that recombination may play in creating or removing sequence variation. During meiosis, the reciprocal recombination (crossing-over) that occurs between homologs to ensure their segregation can also act to increase sequence variation both by creating new combinations of alleles on the

recombinant chromatids and by separating deleterious combinations of mutant alleles. However, both meiotic and mitotic cells are also proficient at gene conversion, which can act to decrease variation by correcting mutant alleles to wild-type or vice versa. In fact, intra- or interchromosomal gene conversion events within tandem or inverted repeat arrays has been proposed as a mechanism for maintaining the homogeneity of sequences within repeat arrays in other organisms, mostly notably the rDNA repeats in *Drosophila* (cf. Dover et al., 1982). So it seems likely that such gene conversion events might act within the human Y chromosome to maintain sequence homogeneity between the repeats of a given Y chromosome.

The Importance of Being Ampliconic, the Functional Significance of These Structures

It remains to be seen whether or not the conversion events observed here reflect the action of a mechanism that functions specifically or preferentially to promote gene conversion within the array. Despite the observation that on average 600 duplicated nucleotides have undergone arm-to-arm gene conversion per transmitted Y chromosome, this only reflects a conversion of some .006% of the ampliconic regions per transmitted Y, a low frequency indeed. Given that most conversion events span only a few hundred base pairs of contiguous sequence, the amount of gene conversion observed here can be explained by only 1–2 gene conversion events in the mitotic or meiotic history of each Y chromosome per generation. Thus, rather than imagining the existence of some system that specifically ensures intrachromosomal recombination on the Y chromosome, it may be more reasonable to propose that such conversions result from the repair of occasional DSBs within the ampliconic cluster in either meiosis or mitosis. Moreover, conversion events within the ampliconic region might be expected to have immediate effects on the transmissibility of a given Y chromosome. Conversion events that "corrected" a recently arising loss-of-function mutation present in one copy of a testis-specific gene back to the wild-type sequence would further ensure the production of normal levels of the protein encoded by these genes during the next round of spermatogenesis. On the other hand, conversion events in which the wild-type copy of the gene was replaced by sequence from the mutant copy (and thus inactivated) might further reduce the ability of that Y chromosome to produce that protein during the next round of spermatogenesis, and thus help to ensure the loss of cells bearing that Y chromosome during the next round of spermatogenesis.

Evidence in support of a role of this type of selection in maintaining the integrity of these ampliconic genes derives from an "Achilles Heel" created by their structure. Palindromes P1and P2 are flanked by a pair of direct repeats (denoted b4 and b2) of a sequence approximately 200 kb in length (Kuroda-Kawaguchi et al. 2001). Exchanges between these repeats, either as intra-chromosomal "loop-out" exchanges or unequal exchanges between sister chromatids, result in Y chromosomes that have lost both palindromic arrays (Kuroda-Kawaguchi et al., 2001; Yen, 2001). The frequency of such exchanges is high enough that the resulting deletions occur at a frequency of about 3×10^{-4}, and such deletions may account for as many as one in 10 of all infertile males (Yen, 2001). A likely component of that sterility is the loss of the DAZ genes, which are located in palindromes P1 and P2 and which, as noted above, seem likely to be required at multiple points in male germ cell development (Reijo et al., 2000). Thus, the loss of the testis-specific genes, or perhaps even a significant reduction in their copy number, is expected to have immediate effects on male fertility, perhaps ending that lineage of Y chromosomes within the same male in which it was created.

Both the acquisition of male-specific and male-essential functions by the Y chromosome, and the unusual structure of the ampliconic elements in which they reside, may provide a powerful selective force for maintaining the Y chromosome in the human population. Indeed, unusual genetic structures are not uncommon for essential elements on the Y chromosomes of other species as well. Consider for example, the satellite DNA-rich lampbrush loop-like structures that comprise the fertility factors of the *Drosophila hydei* Y chromosome (Reugels et al. 2000) or the fertility factors of the *Drosophila melanogaster* Y chromosome in which genes encoding testis-specific forms of dynein and protein phosphatases have been swollen to enormous sizes by the presence of a large number of heterochromatic introns (Carvalho et al., 2001). The presence of unusual DNA structures in regions containing essential fertility genes, in creatures as evolutionarily distant as are humans and flies, makes one wonder whether the unusual DNA structures themselves may also be essential for modulating gene expression during spermatogenesis.

All of this suggests that the obituary for the Y chromosome noted above may have been written rather prematurely. At the very least, the MSY ampliconic regions are far from being a wasteland or a wrecking yard, but rather appear to be highly evolved and complex genetic structures that contain genes essential for male fertility. Moreover, the predominance of genes that are unique to the Y chromosome, and which encode testis-specific functions, within the ampliconic regions provides a powerful selective pressure to maintain these regions, even in the presence of an X chromosomal homolog. Similarly, the presence of a conversion-based system of gene copy "correction" acts to prevent the gradual accumulation of deleterious mutants that might be expected in the absence of meiotic recombination with a homolog. Finally, we note that the authors now use the term MSY to denote a region once referred to as the NRY, for nonrecombining region of the Y chromosome. This is clearly an essential name change, since whatever this region is doing, or not doing, it is very clearly recombining!

Selected Reading

Aitken, R.J., and Marshall Graves, J.A. (2002). Nature *415*, 963.

Carvalho, A.B., Dobo, B.A., Vibranovski, M.D., and Clark, A.G. (2001). Proc. Natl. Acad. Sci. USA *98*, 13225–13230.

Dover, G.A., Strachan, T., Coen, E.S., and Brown, S.D. (1982). Science *218*, 1069.

Kuroda-Kawaguchi, T., Skaletsky, H., Brown, L.G., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., Silber, S., Oates, R., Rozen, S., and Page, D.C. (2001). Nat. Genet. *29*, 279–286.

Lahn, B.T., Pearson, N.M., and Jegalian, K. (2001). Nat. Rev. Genet. *2*, 207–216.

Marshall Graves, J.A. (2000). Biol. Reprod. *63*, 667–676.

Reijo, R.A., Dorfman, D.M., Slee, R., Renshaw, A.A., Loughlin, K.R., Cooke, H., and Page, D.C. (2000). Biol. Reprod. *63*, 1490–1496.

Reugels, A.M., Kurek, R., Lammermann, U., and Bunemann, H. (2000). Genetics *154*, 759–769.

Rozen, S., Skaletsky, H., Marzalek, J.D., Minx, P.J., Waterston, R.H., Wilson, R.K., and Page, D.C. (2003). Nature, *423*, 873–876.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx Holland, P.J., Cordum, S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). Nature, *423*, 825–837.

Yen, P. (2001). Nat. Genet. *29*, 243–244.