

A Large Palindrome With Interchromosomal Gene Duplications in the Pericentromeric Region of the *D. melanogaster* Y Chromosome

María Méndez-Lago,^{†1} Casey M. Bergman,^{†2} Beatriz de Pablos,¹ Alan Tracey,³ Siobhan L. Whitehead,³ and Alfredo Villasante^{*,1}

¹Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Universidad Autónoma de Madrid, Cantoblanco, Madrid, Spain

²Faculty of Life Sciences, University of Manchester, Manchester, United Kingdom

³The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: avillasante@cbm.uam.es.

Associate editor: Jody Hey

Abstract

The non-recombining Y chromosome is expected to degenerate over evolutionary time, however, gene gain is a common feature of Y chromosomes of mammals and *Drosophila*. Here, we report that a large palindrome containing interchromosomal segmental duplications is located in the vicinity of the first amplicon detected in the Y chromosome of *D. melanogaster*. The recent appearance of such amplicons suggests that duplications to the Y chromosome, followed by the amplification of the segmental duplications, are a mechanism for the continuing evolution of *Drosophila* Y chromosomes.

Key words: Y chromosome, palindrome, pericentromeric region, segmental duplications, *D. melanogaster*.

Almost a century ago, Muller (1914) first suggested that the Y chromosome is a highly degenerate X chromosome. Since then, several models have tried to explain how the heteromorphic sex chromosomes have evolved from a pair of autosomes (Ohno 1967; Charlesworth 1978; Rice 1994; Graves 1995). The prevailing theory holds that after the acquisition of a sex-determining gene by one autosome, suppression of recombination leads to its degeneration by the accumulation of deleterious mutations. Although the relative importance of the mechanisms governing this evolutionary transition is still an open issue, the presence of X-degenerated genes on the primate Y chromosome (Skaletsky et al. 2003; Hughes et al. 2010) and Z-degenerated sequences on the avian W chromosome (Fridolfsson et al. 1998) provide empirical evidence for this hypothesis. The three major evolutionary processes that account for the degradation of Y-linked genes are Muller's ratchet, genetic hitchhiking, and background selection (Charlesworth B and Charlesworth D 2000; Bachtrog 2004). Furthermore, the relative contributions of these evolutionary forces over the course of Y chromosome degeneration have recently been formally addressed (Bachtrog 2008).

In *Drosophila melanogaster*, the lack of a male-determining gene on the Y chromosome and the lack of sequence homology, other than the ribosomal DNA, between the X and Y chromosomes have given rise to the hypothesis that the *Drosophila* Y chromosome evolved from a supernumerary B chromosome (Hackstein et al. 1996; Carvalho 2002). The few *D. melanogaster* Y-linked genes so far identified have male-related functions and all originated from duplication of autosomal genes (Kennison 1981; Gepner and

Hays 1993; Carvalho 2002; Koerich et al. 2008). However, the acquisition of autosomal genes is not restricted to non-canonical Y chromosomes; in fact, the deleted in azoospermia gene cluster on the human Y chromosome was the first clear example of an autosomal fertility gene that moved onto a Y chromosome (Saxena et al. 1996).

The sequencing of the human (Skaletsky et al. 2003), chimpanzee (Hughes et al. 2010), and mouse (Alföldi 2008) Y chromosomes led to the discovery of ampliconic regions (made of large repeat units) containing multicopy genes required for male fertility. In primates, these amplicons are principally arrayed as palindromes, and their sequence identity is maintained by frequent gene conversion between repeated regions (Rozen et al. 2003; Hughes et al. 2010). Thus, although these Y chromosomes do not recombine with the X chromosomes, their ampliconic regions are proposed to conserve gene functions over evolutionary time by means of gene conversion (Charlesworth 2003; Rozen et al. 2003). Recently, theoretical and computer simulation studies have shown that low rates of gene conversion can oppose the degeneration of Y-linked genes (Connallon and Clark 2010; Marais et al. 2010).

Assessing whether amplicons and large palindromic regions are a common feature of Y chromosome evolution requires further studies in other nonmammalian species. In this context, it has been shown in *D. melanogaster* that the Y-linked *Su(Ste)* repeats, which affect male fertility, derive from a recent duplication and amplification of the autosomal *SSL* gene (Kalmykova et al. 1997; Gvozdev et al. 2005). Likewise, Russell and Kaiser (1993) found that a Y chromosome pericentromeric region also derives from a recent

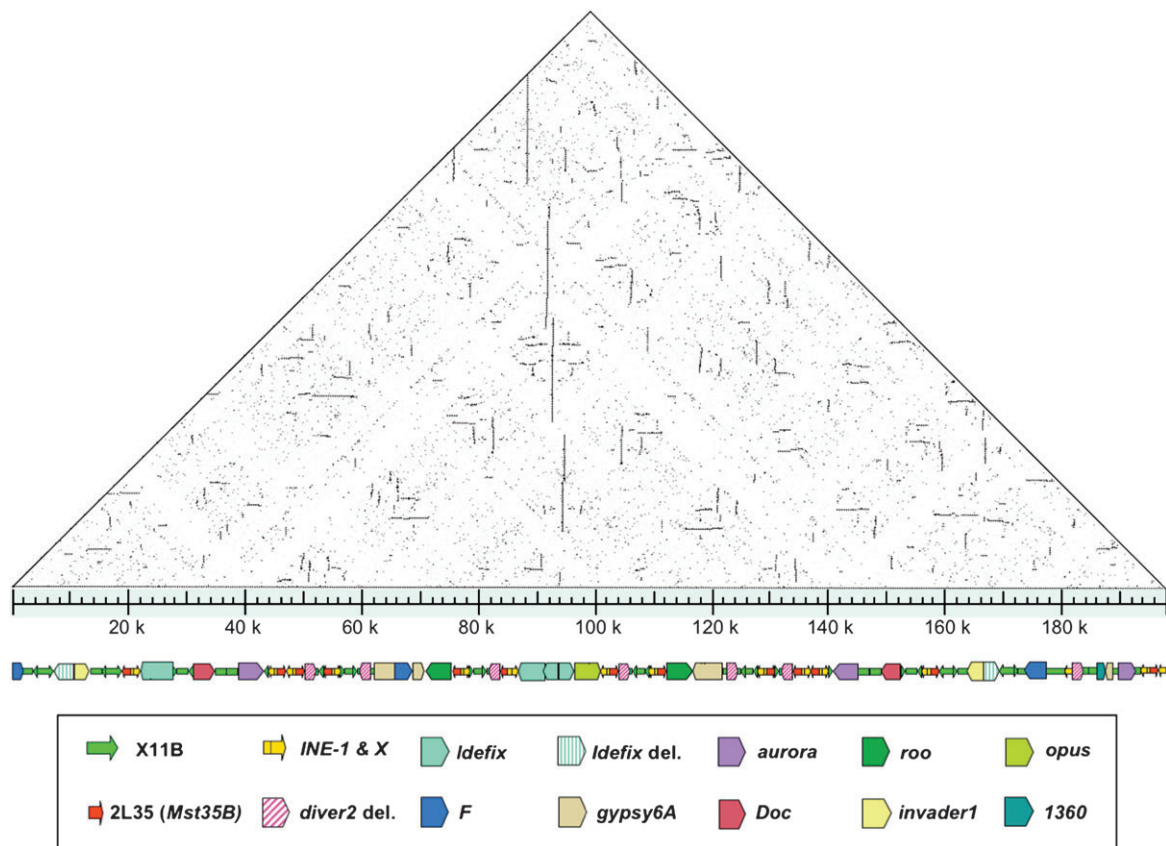


Fig. 1. Schematic representation of the BACR07N15 sequence. The sequence organization of the large palindrome is represented by the horizontal diagram at the bottom. Sequence similarity within BACR07N15 is depicted above as a triangular dot plot (made with the program “dotter”) in which the BACR07N15 sequence is compared with itself. Direct repeats appear as horizontal lines and inverted repeats as vertical lines.

segmental duplication of the *Mst77F* locus and subsequent amplification. Recently, Carvalho et al. have shown that this ampliconic region contains at least 18 members of the *Mst77Y* gene family, 8 of which are potentially functional (Krsticevic et al. 2010). However, previous analyses of the assembled genome sequence have not found any large palindromes on the Y chromosome or any other chromosome (Carlson JW, Celniker SE, personal communication).

To investigate the presence of additional pericentromeric amplicons on the *D. melanogaster* Y chromosome and their potential palindromic organization, we isolated several pericentromeric bacterial artificial chromosomes (BACs) using heterochromatin P-element insertions (Yan et al. 2002), given that the *Mst77Y* amplicon had previously impeded extension of the Y centromeric region by chromosome walking (Abad et al. 2004). Using the flanking sequence of B783.2, a P-element inserted at the h17–18 region, four BACs were detected by an “in silico” search (supplementary fig. S1A, Supplementary Material online). DNA restriction analysis of the biggest ones, BACR36F19 and BACR07N15, showed that they derived from the same region, and fluorescence in situ hybridization mapping on mitotic chromosomes confirmed their pericentromeric origin (supplementary fig. S1B–C, Supplementary Material online). To investigate the chromosomal organization of this region, the clone BACR07N15 (~200 kb) was sequenced

(GenBank: CU457433) and compared with the *D. melanogaster* genome (Release 5.1). Sequence analysis showed that the region contained in this BAC is not part of the publicly available Y chromosome sequences. In addition, dotplot analysis revealed the presence of a large palindrome (fig. 1).

We reconstructed the history of this region and found that this large palindrome arose by a series of distinct molecular processes (fig. 2): 1) transposition to the Y of two interchromosomal duplications, one from the cytological band 35B (chromosome arm 2L) and the other from 11B (chromosome X), 2) intrachromosomal amplification of a fragment containing these segmental duplications, 3) small inversions and deletions through the entire region, 4) insertion of multiple transposable element (TE) sequences, and 5) a large tandem duplication resulting from a single recombination event within two *Idefix* long terminal repeats, followed by a large inversion resulting from intrachromatid pairing of two *Idefix* elements, double-strand breaks inside these elements and nonhomologous end-joining between the broken elements. This reconstruction is supported by the detailed annotation of TEs in BACR07N15. The majority of these TEs are from the same family and in the same relative inverted positions in both arms of the palindrome (fig. 1), and thus, we assume their insertion predates the origin of the palindrome since the

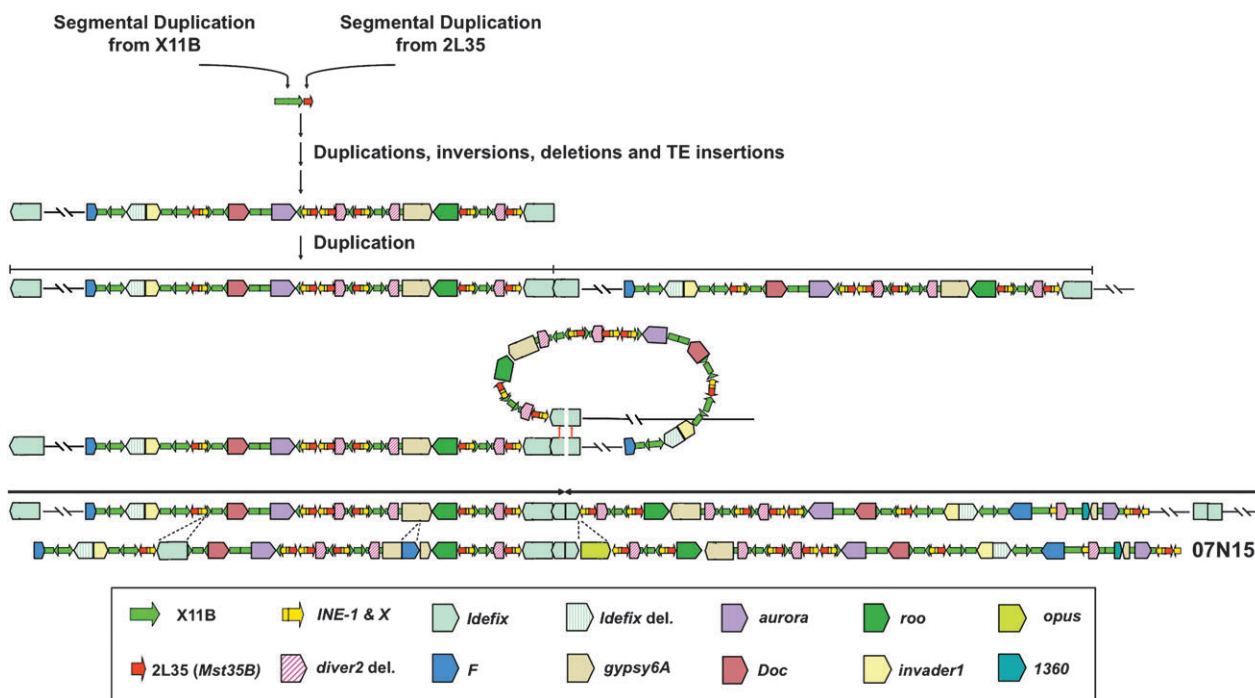


Fig. 2. Proposed model for palindrome formation. The evolutionary processes hypothesized to be involved in the generation of the palindrome are a large tandem duplication followed by a large inversion. In the middle of the palindrome, there is an inverted duplication of an *Idefix* element with a small portion of sequence missing at the breakage-fusion junction. Because a double-strand break repaired with concomitant loss of sequence information is characteristic of nonhomologous end-joining, this is the mechanism proposed for the origin of the palindrome. The 12 TE families and the 2 segmental duplications are indicated.

alternative model that these TE insertions arose in one arm and were copied to the other arm by gene conversion is less parsimonious. Under this model, insertions that are unique to only one or the other arm of the palindrome are assumed to have inserted after the palindrome formation.

To provide an estimate of the sequence divergence between arms of the palindrome, we compared the 72 kb of unambiguously alignable DNA from each arm present in BACR07N15. This analysis provided an overall divergence between arms of 0.0036 sub/site using Kimura's 2-parameter method. This arm-to-arm nucleotide sequence divergence is greater than in human Y palindromes (<0.06%, Rozen et al. 2003) but similar to the 0.5% arm-to-arm divergence observed for some chimpanzee Y palindromes (Hughes et al. 2010). The arms of the BACR07N15 palindrome show high sequence similarity either because of recent origin and/or ongoing gene conversion. Assuming that TE insertions present in both arms predate the palindrome formation, the age of TE insertions that are present in either one or both arms of the palindrome provides an independent method to place lower and upper bounds on the age of the palindrome (supplementary results and fig. S2, Supplementary Material online). This analysis, together with analysis of the CG12717/CG41423 gene family (see below), favors the interpretation of a recent origin for the palindrome as the primary reason why arms of the palindrome exhibit high sequence similarity. If this interpretation is correct and assuming a molecular clock of 0.011 subs/Ma (Tamura et al. 2004), we obtain an age estimate of ~0.33 Ma for the origin of the palindrome. This result is consistent with the palin-

drome being *D. melanogaster* specific because the divergence of *D. melanogaster* from its sister species *D. simulans* occurred ~5.5 Ma (Tamura et al. 2004). We note, however, that despite providing evidence for a recent origin, our data cannot conclusively rule out the effects of subsequent gene conversion between the palindrome arms.

We next addressed the extent of the palindrome by further physical mapping using probes derived from BACR07N15. Sequence comparison of the duplications and their ancestral loci suggests that the segmental duplication 2L35B derived from a genomic fragment that spanned at least the first intron and the second exon of *Mst35Ba* (one of the two protamine genes in tandem) (supplementary fig. S3A, Supplementary Material online), whereas the segmental duplication X11B derived from a fragment encompassing the whole CG12717 gene (encoding a SUMO protease) and parts of the neighbor genes *ade5* and CG3812 (supplementary fig. S3B, Supplementary Material online). A search for other CG12717-related gene models in the *D. melanogaster* genome revealed the presence of CG41423, which is located in the unmapped whole-genome shotgun scaffold AABU01001566. Because the X11B derived CG12717 fragments in BACR07N15 are highly similar to CG41423 (see supplementary fig. S3C, Supplementary Material online), we screened the BAC library RPCI-98 with a probe from the *Mst35Y* intron (supplementary methods, Supplementary Material online) to see if the AABU01001566 scaffold belongs to another region of the palindrome identified in BACR07N15. Restriction analyses of the seven strongly hybridizing BACs allowed us to build

a contig of about 500 kb with a palindromic region of at least 350 kb (supplementary fig. S4A, Supplementary Material online). By means of polymerase chain reaction amplification, an entire CG41423 gene was isolated from each of the BACs that partially overlap BACR07N15 and extend in both directions (supplementary fig. S4B, Supplementary Material online). Moreover, an “in silico” search detected two BACs from library DM1 (DME1-001M04 and DME1-007E02) that extend the palindrome to the *Mst77Y* amplicon (Krsticevic et al. 2010). As the average insert size of this library is 100 kb, the palindrome identified here could be located at less than 100 kb of the *Mst77Y* amplicon.

Finally, we addressed the potential functionality of gene sequences in the BACR07N15 palindrome by molecular evolutionary analysis. BlastN searches in the genome of sequenced species of the melanogaster subgroup have failed to identify any CG41423-like gene, other than the one from the X chromosome. This fact and the phylogenetic position of CG41423 (supplementary fig. S5, Supplementary Material online) strongly suggests that the initial duplication from the X to the Y chromosome occurred after the split of the *D. melanogaster* and *D. simulans* lineages, supporting the inference of recent origin in *D. melanogaster* for the palindrome itself. CG41423 has evolved under purifying selection ($d_N/d_S = 0.3837$) suggesting recent functional constraint. Although CG41423 transcripts have been detected by the modENCODE project, no cDNA has yet been found in expressed sequence tag databases. The potential SUMO protease encoded by CG41423 could be implicated in spermiogenesis-related sumoylation events.

Evolutionary analysis of the decayed *Mst35Y* genes shows, as happens in primate protamine genes (Retief and Dixon 1993), that their exons are more variable than their introns. Although the *Mst35Y* genes in BACR07N15 appear nonfunctional, it remains possible that the missing portion of the palindrome, or a flanking region might contain a functional *Mst35Y*, as has been shown for *Mst77Y* (Krsticevic et al. 2010), which would act in redundancy to *Mst35Ba* and *Mst35Bb*. This interpretation is consistent with the fact that the *Mst35Y* genes in BACR07N15 show evidence of functional constraint ($d_N/d_S = 0.0114$). The existence of a functional protamine gene on the Y chromosome would also explain why *D. melanogaster* males homozygous for a deletion of both protamine genes, *Mst35Ba* and *Mst35Bb*, are fertile (Rathke et al. 2010). Lastly, we note that the two segmental duplications found at the pericentromeric region of the *D. melanogaster* Y (Krsticevic et al. 2010; this work) show functional as well as structural (see above) coherence because *Mst35B* and *Mst77F* encoded for a protamine and a sperm-specific linker histone-like protein, respectively (Raja and Renkawitz-Pohl 2005).

We emphasize that the palindromic structure identified here in the *Drosophila* Y chromosome differs from its mammalian counterparts in two essential aspects. First, our data provide no direct evidence for gene conversion between arms of the palindrome. Although the lack of meiotic crossing in *D. melanogaster* males (Morgan 1912) would seem to preclude the possibility of gene conversion between palin-

dromic regions on the Y chromosome, mitotic recombination can occur in *D. melanogaster* males and lead to gene conversion (Preston and Engels 1996). Additionally, there is molecular evidence of intrachromosomal gene conversion among copies of the Y-linked *Su(Ste)* locus in *D. melanogaster* (Balakireva et al. 1992; McKee and Satter 1996) and duplicates of the Y-linked *kl2* gene in *D. simulans* complex (Kopp et al. 2006). Second, whereas the human Y palindromes contain functional genes essential for male fertility, the selective advantage for gene conversion to preserve function on the *Drosophila* Y palindrome is likely to be small because it has not been shown to contain essential genes. Nevertheless, the discovery of palindromic structures on the *Drosophila* Y chromosome that contain genes provides the first step toward establishing the existence of functional palindromes in nonmammalian species.

Supplementary Material

Supplementary results, methods, references, and figures S1–S5 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We would like to thank Keith Maggert for providing the flanking sequence of the P-element insertion B783.2 and Patrizio Dimitri for confirming the chromosomal location of BACR07N15. We are also grateful to Jane Rogers for her support. This work was supported by the Ministerio de Ciencia e Innovación (BFU2008-02947-C02-01/BMC to A.V.); an institutional grant from the Fundación Ramón Areces to the Centro de Biología Molecular “Severo Ochoa”; the Natural Environment Research Council (NERC NE/G000158/1 to C.M.B.); and The Wellcome Trust. The GenBank accession number for the sequence reported in this paper is CU457433.

References

- Abad JP, de Pablos B, Agudo M, Molina I, Giovinnazo G, Martín-Gallardo A, Villasante A. 2004. Genomic and cytological analysis of the Y chromosome of *Drosophila melanogaster*: telomere-derived sequences at internal regions. *Chromosoma* 113:295–304.
- Alföldi JE. 2008. Sequence of the mouse Y chromosome [dissertation]. Cambridge, Massachusetts: Massachusetts Institute of Technology.
- Bachtrog D. 2004. Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nat Genet*. 36:518–522.
- Bachtrog D. 2008. The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* 179:1513–1525.
- Balakireva MD, Shevelev YY, Nurminsky DI, Livak KJ, Gvozdev VA. 1992. Structural organization and diversification of Y-linked sequences comprising *Su(Ste)* gene in *Drosophila melanogaster*. *Nucleic Acids Res*. 20:3731–3736.
- Carvalho AB. 2002. Origin and evolution of the *Drosophila* Y chromosome. *Curr Opin Genet Dev*. 12:664–668.
- Charlesworth B. 1978. Model for evolution of Y chromosomes and dosage compensation. *Proc Natl Acad Sci U S A*. 75:5618–5622.
- Charlesworth B. 2003. The organization and evolution of the human Y chromosome. *Genome Biol*. 4:226.

- Charlesworth B, Charlesworth D. 2000. The degeneration of Y chromosomes. *Philos Trans R Soc Lond B Biol Sci.* 355:1563–1572.
- Connallon T, Clark AG. 2010. Gene duplication, gene conversion and the evolution of the Y chromosome. *Genetics* 186:277–286.
- Fridolfsson AK, Cheng H, Copeland NG, Jenkins NA, Liu HC, Raudsepp T, Woodage T, Chowdhary B, Halverson J, Ellegren H. 1998. Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc Natl Acad Sci U S A.* 95: 8147–8152.
- Gepner J, Hays TS. 1993. A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. *Proc Natl Acad Sci U S A.* 90:11132–11136.
- Graves JAM. 1995. The origin and function of the mammalian Y chromosome and Y-borne genes—an evolving understanding. *Bioessays* 17:311–320.
- Gvozdev VA, Kogan GL, Usakin LA. 2005. The Y chromosome as a target for acquired and amplified genetic material in evolution. *Bioessays* 27:1256–1262.
- Hackstein JHP, Hochstenbach R, Hauschteck-Jugen E, Beukeboom LW. 1996. Is the Y chromosome of *Drosophila* an evolved supernumerary chromosome? *BioEssays.* 18:317–323.
- Hughes JF, Skaletsky H, Pyntikova T, et al. (14 co-authors). 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463:536–539.
- Kalmykova AI, Shevelyov YY, Dobritsa AA, Gvozdev VA. 1997. Acquisition and amplification of a testis-expressed autosomal gene, SSL, by the *Drosophila* Y chromosome. *Proc Natl Acad Sci U S A.* 94:6297–6302.
- Kennison JA. 1981. The genetic and cytological organization of the Y chromosome of *Drosophila melanogaster*. *Genetics* 98: 529–548.
- Koerich LB, Wang X, Clark AG, Carvalho AB. 2008. Low conservation of gene content in the *Drosophila* Y chromosome. *Nature* 456:949–951.
- Kopp A, Frank AK, Barmina O. 2006. Interspecific divergence, intrachromosomal recombination, and phylogenetic utility of Y-chromosomal genes in *Drosophila*. *Mol Phylogenet Evol.* 38: 731–741.
- Krsticevic FJ, Santos HL, Januário S, Schrago CG, Carvalho AB. 2010. Functional copies of the Mst77F gene on the Y chromosome of *Drosophila melanogaster*. *Genetics* 184:295–307.
- Marais GA, Campos PR, Gordo I. 2010. Can intra-Y gene conversion oppose the degeneration of the human Y chromosome? A simulation study. *Genome Biol Evol.* 2:347–57.
- McKee BD, Satter MT. 1996. Structure of the Y chromosomal Su(Ste) locus in *Drosophila melanogaster* and evidence for localized recombination among repeats. *Genetics* 142:149–161.
- Morgan TH. 1912. Complete linkage in the second chromosome of the male of *Drosophila*. *Science* 36:719–720.
- Muller HJ. 1914. A gene for the fourth chromosome of *Drosophila*. *J Exp Zool.* 17:325–336.
- Ohno S. 1967. Sex chromosomes and sex linked genes. New York: Springer-Verlag.
- Preston CR, Engels WR. 1996. P-element-induced male recombination and gene conversion in *Drosophila*. *Genetics* 144:1611–1622.
- Raja SJ, Renkawitz-Pohl R. 2005. Replacement by *Drosophila melanogaster* protamines and Mst77F of histones during chromatin condensation in late spermatids and role of sesame in the removal of these proteins from the male pronucleus. *Mol Cell Biol.* 25:6165–6177.
- Rathke C, Barckmann B, Burkhard S, Raja SJ, Roote J, Renkawitz-Pohl R. 2010. Distinct functions of Mst77F and protamines in nuclear shaping and chromatin condensation during *Drosophila* spermiogenesis. *Eur J Cell Biol.* 89:326–338.
- Retief JD, Dixon GH. 1993. Evolution of pro-protamine P2 genes in primates. *Eur J Biochem.* 214:609–615.
- Rice WR. 1994. Degeneration of a nonrecombining chromosome. *Science* 263:230–232.
- Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423:873–876.
- Russell SRH, Kaiser K. 1993. *Drosophila melanogaster* male germ line-specific transcripts with autosomal and Y-linked genes. *Genetics* 134:293–308.
- Saxena R, Brown LG, Hawkins T, et al. (11 co-authors). 1996. The DAZ gene cluster on the human Y chromosome arose from an autosomal gene that was transposed, repeatedly amplified and pruned. *Nat Genet.* 14:292–299.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, et al. (40 co-authors). 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Tamura K, Subramanian S, Kumar S. 2004. Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol.* 21:36–44.
- Yan CM, Dobie KW, Le HD, Konev AY, Karpen GH. 2002. Efficient recovery of centric heterochromatin P-element insertions in *Drosophila melanogaster*. *Genetics* 161:217–229.