nature genetics

Adaptation shapes patterns of genome evolution on sexual and asexual chromosomes in *Drosophila*

Doris Bachtrog

What advantage might sexual recombination confer? Population genetics theory predicts that asexual genomes are less efficient at eliminating deleterious mutations and incorporating beneficial alleles[1,2]. Here, I compare patterns of genome evolution in a 40-kb gene-rich region on homologous neo-sex chromosomes of *Drosophila miranda*[3]. Genes on the non-recombining neo-Y show various signs of degeneration, including transposable-element insertions, frameshift mutations and a higher rate of amino-acid substitution. In contrast, loci on the recombining neo-X show intact open reading frames and generally low rates of amino-acid substitution. One exceptional gene on the neo-X shows evidence for adaptive protein evolution, affecting patterns of variability at neighboring regions along the chromosome. These findings illustrate the limits to natural selection in an asexual genome. Deleterious mutations, including repetitive DNA, accumulate on a non-recombining chromosome, whereas rapid protein evolution due to positive selection is confined to the recombining homolog.

One challenge of evolutionary biology is to elucidate the adaptive significance of sexual recombination[1,4]. Despite the automatic reproductive advantage of asexual reproduction, most higher organisms reproduce sexually[5]. Asexual species seem to persist for much shorter evolutionary periods than do their sexual relatives[6]. Asexual genomes are prone to degeneration; an example is the evolutionary fate of Y chromosomes[4,7]. The X and Y chromosome descended from originally homologous autosomes[7,8]; the almost complete erosion of genetic activity on the Y chromosome is a direct consequence of its lack of recombination[7].

On ancient Y chromosomes, like the human Y, only few functional genes persist and a large fraction consists of repetitive 'junk' DNA. To study the processes underlying the degeneration of non-recombining genomes, I used the recently formed neo-sex chromosomes of *D. miranda*. Due to the fusion of an autosome to the Y chromosome, a formerly autosomal chromosome arm is transmitted with the true Y (the neo-Y chromosome). Because male *Drosophila* lack recombination, the neo-Y is completely sheltered from recombination and is thus transmitted clonally. Indeed, despite its recent origin (~1 million years[3]; after the split from its close relative *D. pseudoobscura*), the neo-Y already shows substantial levels of degeneration[3,9,10].

Evolutionary theory predicts that a non-recombining genome should have a lower effective population size, N_e[1,11]. Patterns of nucleotide polymorphism suggest that N_e of the neo-Y chromosome of *D. miranda* is about 30 times smaller than that of the neo-X[3,12]. The reduction in N_e reduces the efficacy of natural selection[1]; the rate of accumulation of slightly deleterious mutations should be higher on a non-recombining chromosome[2] and the rate of adaptation should be lower[13].

To study the molecular and population genetic forces involved in the evolution of recombining versus non-recombining genomes, I isolated large genomic fragments located on the neo-sex chromosomes[14] and the orthologous (autosomal) *D. pseudoobscura* sequence, which permits the inference of lineage-specific changes on the neo-X and neo-Y. The region was annotated using the *D. melanogaster* genome as a reference[15]; it contains the gene *exuperantia1* and six additional genes predicted from the *D. melanogaster* genomic sequence (**Fig. 1**).

All seven genes have intact open reading frames on the neo-X chromosome of *D. miranda*, and patterns of molecular evolution indicate that all genes from the neo-X evolve under selective constraint (**Table 1**). In particular, the rate of non-synonymous substitutions is low relative to the rate for synonymous substitutions at most genes, as expected for loci evolving under strong functional constraints[16].

In contrast, the neo-Y-linked region shows various signs of degeneration. There are four large insertions on the neo-Y chromosome, representing retrotransposons (**Fig. 1a**). Indeed, almost half of the neo-Y-linked sequence is derived from transposable-element DNA. *worf* (4.2 kb) is inserted in the first intron of the gene *CG9025*, *kirk* (4.2 kb) is integrated in the first intron of the gene *CG16799* and *spock* (5.0 kb) and *picard* (4.0 kb) are inserted into intergenic regions (**Fig. 1a**). All four transposable elements are in excess copy number on the neo-Y of *D. miranda* (ref. 17 and D.B., unpublished data). Whereas most transposons segregate at low frequencies in natural populations of *Drosophila* (as expected for mildly deleterious mutations[18]), the four transposable elements investigated are fixed at their genomic location in the sample investigated (ref. 17 and D.B., unpublished data). This reflects the inability of natural selection to prevent the accumulation of repetitive DNA on the non-recombining neo-Y chromosome[10,19].

Institute of Cell, Animal and Population Biology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. e-mail: Doris.Bachtrog@ed.ac.uk

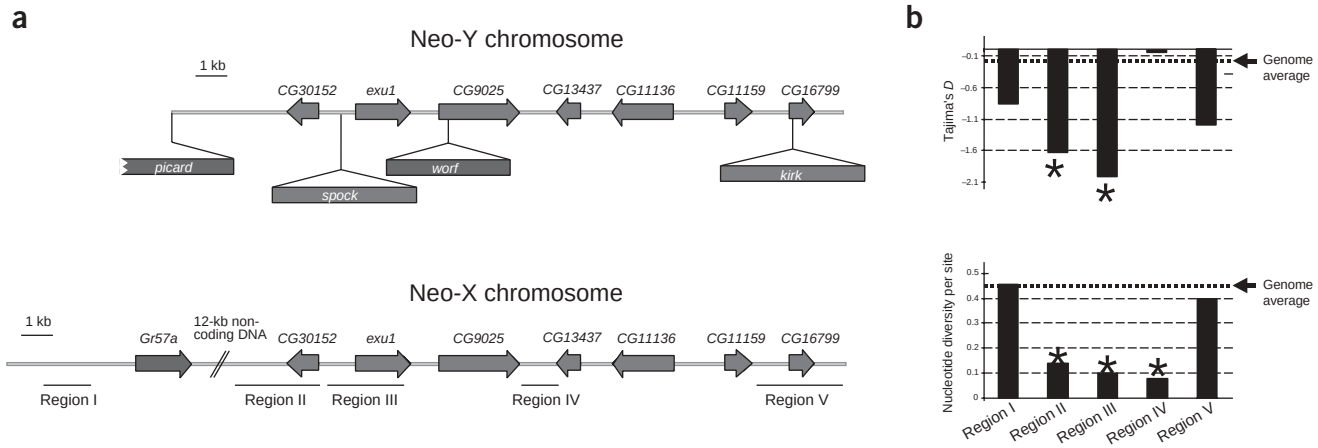

Figure 1 Adaptation and degeneration on the neo-sex chromosomes. (**a**) Organization of the genomic region investigated on the neo-X and the neo-Y chromosomes of *D. miranda*. Approximately 40 kb from the neo-X chromosome and 45 kb from the neo-Y chromosome were analyzed. The overlapping region between the neo-X and neo-Y was also obtained from *D. pseudoobscura* (~21 kb), where it is autosomal. The gene orders in the region analyzed are identical to those in *D. melanogaster*[15]. The neo-Y-linked region contains the insertion of four transposable elements, called *picard*, *worf*, *spock* and *kirk* (ref. 17 and D.B., unpublished data). Fragments I–V correspond to the genomic regions for which polymorphism data was collected for the neo-X chromosome. (**b**) Levels of nucleotide diversity (π, average number of pairwise differences per site) and Tajima's *D* statistics[24] (a measurement of the skew in the frequency spectrum of polymorphisms) at the five fragments sequenced in twelve *D. miranda* strains[3]. Asterisks indicate significant reductions in nucleotide diversity by the HKA test or values of Tajima's *D* that are inconsistent with a model of neutral evolution at the 5% significance level.

Further, some of the neo-Y-linked genes are clearly non-functional (**Fig. 2**). The neo-Y allele of *CG13437* contains a 299-bp deletion, eliminating all of exon 1 (including the start codon) and part of exon 2. *CG9025*, a gene with homologous sequences in yeast, *Caenorhabditis elegans*, mouse and humans, shows a 21-bp in-frame deletion in exon 1, the insertion of the *worf* transposable element in intron 1 and a 493-bp deletion eliminating most of exon 2. The neo-Y-linked copy of *CG30152* has a 12-bp deletion that eliminates the splicing site between intron 1 and exon 2 and results in a frameshift mutation. The remaining four genes on the neo-Y chromosome appear to be translatable into functional proteins.

Genes on the neo-Y also show generally higher rates of amino-acid substitution compared with the neo-X-linked copies (**Table 1** and **Fig. 2**). This is consistent with less effective purifying selection against deleterious amino-acid mutations on the non-recombining neo-Y chromosome[3]. A similar pattern was observed for two additional protein-coding genes on the neo-Y of *D. miranda*[3] (**Fig. 2**). As the neo-Y is confined to males, natural selection should not promote the conservation of genes with a female-specific role on this chromosome. None of the genes investigated, however, have an obvious female-limited function (that is, they are not predominantly expressed in ovaries; **Table 2**).

Among the neo-X-linked genes, only *exu1* shows a higher rate of replacement substitution than its neo-Y homolog. There are 21 amino-acid substitutions on the branch of the phylogeny leading to the neo-X, but only 9 occur on the neo-Y branch (**Fig. 2**). The higher rate of protein evolution on the neo-X chromosome could be caused either by positive selection or relaxed selective constraint. To distinguish between these hypotheses, I sequenced five fragments located on the neo-X in twelve wild-derived strains of *D. miranda* (a total of 11.2 kb per individual, including the entire coding sequence of *exu1*; **Fig. 1a**). The mode of selection can be inferred by comparing genetic variation within species with fixed differences between species for amino-acid and silent substitutions[20]. The gene *exu1* shows a significant excess of amino-acid substitutions versus polymorphism on the neo-X (38 versus 0) compared with silent substitutions (25 versus 4), using *D. pseudoobscura* as an outgroup species ($P = 0.03$). This suggests that positive selection has driven the rapid protein evolution of *exu1* on the neo-X chromosome.

If the fixations of some amino-acid variants in *exu1* were recent, traces of associated selective sweeps should be detectable in patterns of nucleotide variation at linked sites. One signature of selective sweeps is that variability is lower than would be expected based on nucleotide divergence[21,22]. Specifically, a valley in levels of polymorphisms is expected surrounding the target of positive selection[21,22].

Table 1 Patterns of molecular evolution at the genes investigated on the neo-X and neo-Y chromosomes

	Number of codons	Neo-X chromosome			Neo-Y chromosome		
		d_N	d_S	d_N/d_S	d_N	d_S	d_N/d_S
CG30152	213	0	0.95	0	1.57	4.01	0.39
CG11136	808	0	2.79	0	0.96	2.83	0.34
CG13437	149	0	2.02	0	0.56	3.38	0.17
CG9025	329	0	1.83	0	1.10	0.66	1.66
exu1	497	2.01	2.99	0.67	0.82	1.33	0.61
CG11159	146	0	0	0	0.28	2.72	0.10
CG16799	165	0	3.55	0	0.83	2.28	0.37
Average*		0.43	2.35	0.14	0.93	2.30	0.57
		(0)	(2.18)	(0)	(0.96)	(2.56)	(0.56)

*The values in parentheses are the weighted averages excluding *exu1*.

d_N and d_S are the branch-specific rates of amino-acid substitutions and synonymous substitutions, respectively, given as percentages.

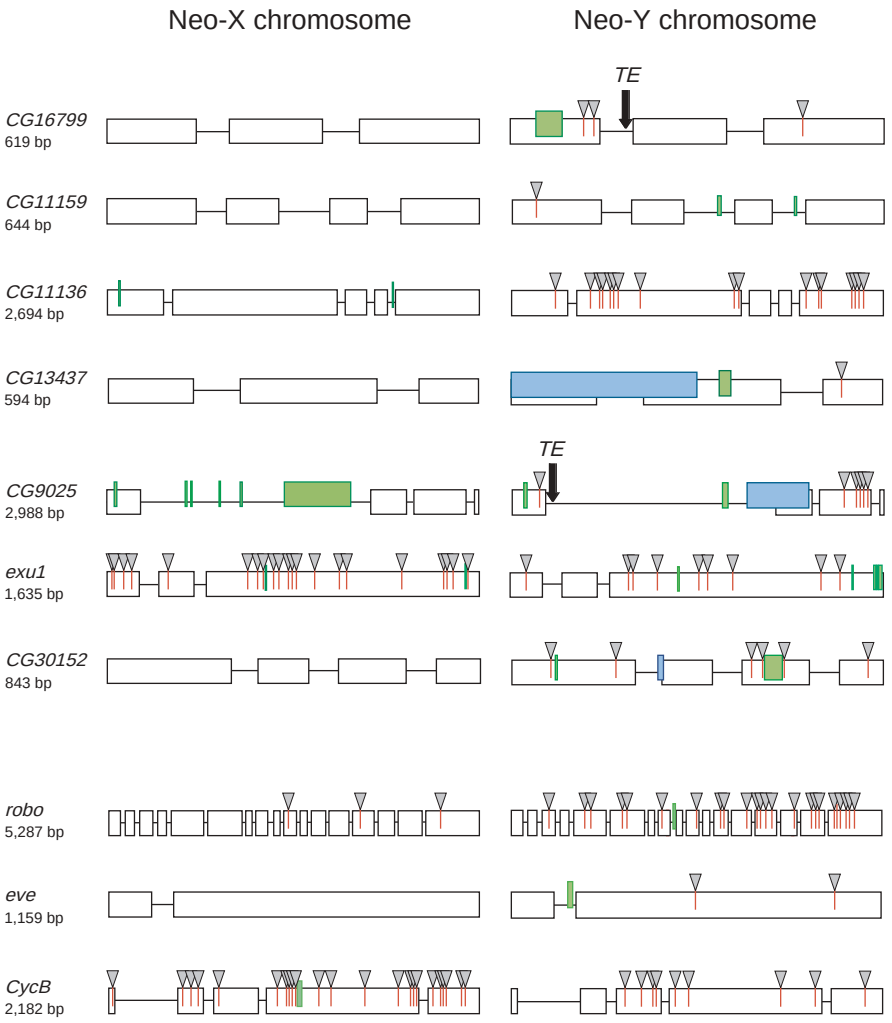

Figure 2 Illustration of amino-acid and insertion-deletion changes in the neo-X- and neo-Y-linked genes. Amino-acid changes are marked as triangles, green boxes indicate in-frame insertions and deletions and blue boxes show frameshift mutations. Mutations were assigned to evolutionary branches using the *D. pseudoobscura* sequence as an outgroup. The genes *CG13437*, *CG9025* and *CG30152* show clear signs of degeneration on the neo-Y. *exu1*, in contrast, shows evidence for adaptive protein evolution on the neo-X chromosome. For comparison, data on three genes previously studied in *D. miranda* are included (*roundabout*, *even-skipped* and *CycB*; ref. 3). *CycB* has also undergone recent adaptive evolution on the neo-X.

The level of silent variation at *exu1* (region III) is 0.10%, about one eighth of the average value of variation observed at three other neo-X-linked loci[3] (**Table 3** and **Fig. 1b**). This reduction in variation is highly significant by the HKA test[21] ($P < 0.01$), using *D. pseudoobscura* as an outgroup. Neighboring fragments also show less polymorphism than expected (region II and IV, $P < 0.05$), whereas more distant genomic fragments show expected levels of variability (region I and V, $P > 0.1$).

Another signature of a selective sweep is that more low-frequency alleles should be observed than expected in a random-mating population with no selection[23]. To test this prediction, I used Tajima's D[23] as a summary statistic. The value of D observed for region III is –1.98, indicating a marked excess of rare alleles ($P < 0.01$). The adjacent fragment II also shows a significantly negative D value (–1.59, $P < 0.05$; **Table 3**), whereas fragments farther away show values of D consistent with neutral evolution. Together, the significant excess in amino-acid divergence, the reduction in neutral polymorphism around *exu1* and the skew in the frequency spectrum of neutral mutations suggest that recent positive selection has driven the rapid evolution of *exu1* on the neo-X chromosome of *D. miranda*.

The gene *exu1* is the second that has been shown to evolve under positive selection on the neo-X of *D. miranda*; like *exu1*, the neo-X-linked copy of the gene encoding cyclin B (*CycB*) shows a high rate of protein-sequence evolution that is driven by positive selection[3] (**Fig. 2**). This suggests that more rapid evolution on the recombining neo-X might be a common phenomenon, potentially involving hundreds of genes. Given the much lower rate of adaptation in an asexual genome, this could result in adaptive downregulation of maladapted genes on the Y chromosome[13].

Sexual recombination is predicted to increase the efficacy of natural selection by allowing the more efficient removal of deleterious

Table 2 Molecular function of the genes investigated on the neo-sex chromosomes

	Biological or molecular function	EST matches
CG30152	unknown	embryo, larvae, pupae
CG11136	ribonuclease inhibitor	none
CG13437	unknown	none
CG9025	unknown	embryo, larvae, pupae, adult heads
exu1	early development, male-female fertility	ovary, testis, embryo
CG11159	lysozyme-like	none
CG16799	lysozyme-like, defense-immunity protein	adult head
eve	early development	embryo
robo	axon guidance	embryo
CycB	protein kinase regulator, cytokinesis, male-female fertility	ovary, testis, embryo, larvae, adult

Table 3 Positive selection at neo-X chromosomal loci

	Region I	Region II	Region III	Region IV	Region V
Number of chromosomes sequenced	12	12	12	12	12
Length of region sequenced	1,543 bp	2,877 bp	2,526 bp	1,302 bp	2,891 bp
Silent sites	1,543 bp	2,416 bp	1,377 bp	1,302 bp	2,442 bp
Silent divergence from *D. pseudoobscura*	2.1%	3.4%	4.2%	2.0%	3.0%
Number of segregating variants[a]	28 (24)	20 (18)	10 (8)	4 (3)	54 (44)
Number of singleton variants[a]	17 (16)	15 (13)	9 (8)	1 (1)	32 (28)
Nucleotide diversity per site (π)	0.44%	0.16% *	0.10% **	0.08% *	0.40%
Theta per site (θ)	0.54%	0.25%	0.19%	0.08%	0.56%
Tajima's *D*	–0.83	–1.59 *	–1.98 **	– 0.03	–1.19

[a]The values in parentheses are the number of variants excluding insertion-deletion polymorphism. *$P < 0.05$, **$P < 0.01$ by the HKA test or based on 10,000 neutral coalescent simulations with no recombination.

mutations and incorporation of beneficial alleles. Artificial selection experiments in *Drosophila* have shown that recombination can prevent the degeneration of a chromosome[24] and increase the rate of adaptation[25] in experimental populations. Here, I show that both phenomena occur in nature. The observed accumulation of repetitive DNA and frameshift mutations on the neo-Y (and their absence on the neo-X) shows the limits of natural selection against deleterious mutations on a non-recombining chromosome. In addition, eight of the ten protein-coding loci have a higher rate of amino-acid evolution on the neo-Y chromosome. Although a higher rate of protein evolution on the neo-Y may also be the result of positive selection, this is unlikely for two reasons. First, the reduced variability of the neo-Y at silent sites[3,12] means that the effective population size, and therefore the efficacy for both purifying and positive selection, will be lower (as generally expected for non-recombining genomes[1,2]). Second, the variance in the number of amino-acid substitutions among genes on the neo-Y is much smaller than that among genes on the neo-X (**Fig. 2**). Amino-acid substitutions occur uniformly among neo-Y-linked genes ($\chi^2 = 7.2$; $P > 0.5$), as expected if all loci on the neo-Y are suffering from lower effectiveness of natural selection (the same is true for presumably neutral silent substitutions: $\chi^2 = 11.4$; $P = 0.25$). In contrast, there is significant heterogeneity in the rate of amino-acid evolution among genes on the neo-X ($\chi^2 = 139.1$; $P < 0.001$), yet no heterogeneity is observed for silent substitutions ($\chi^2 = 8.7$; $P = 0.5$). This strongly suggests that positive selection is limited to only a fraction of genes on the neo-X. *CycB* and *exu1* show clear signatures of positive selection on the recombining neo-X, corroborating the importance of sexual recombination to increase the rate of adaptation. The greater efficiency with which natural selection can both incorporate beneficial mutations and purge deleterious alleles give a population-level benefit to sexual recombination.

What might the nature of the selective pressure acting on *exu1* and *CycB* be? Notably, apart from their well characterized function in development (*exu1*) and cell-cycle regulation (*CycB*), both genes are also expressed in testes and ovaries, and male and female mutants are sterile. The fact that all neo-Y genes are male-limited raises the possibility that the neo-Y and neo-X versions of *exu1* and *CycB* are undergoing adaptive specialization for male and female functions. Indeed, all other genes investigated are not predominantly expressed in testes and ovaries (that is, they do not seem to be involved in sex-specific functions; **Table 2**). In fact, sex-related genes are non-randomly distributed among chromosomes in *D. melanogaster*; genes showing female-biased expression are more abundant (and genes showing male-biased expression are relatively infrequent) on the X[26], whereas the Y chromosome contains mainly genes with male-specific functions[27]. The formation of a non-recombining sex-chromosome pair might be viewed as a genome duplication event that may allow sex-related genes to subfunctionalize into male- and female-specific roles[28].

METHODS

Cloning, localization and sequence analysis of the genomic regions of the neo-X and the neo-Y chromosomes of *D. miranda*. I screened a genomic library of *D. miranda*[3] for the neo-Y-linked and the neo-X-linked copy of *exu1*. I carried out library screening and standard molecular techniques as described[17]. I prepared DNA from the isolated phages and subcloned it with the pZero2.1 vector (Invitrogen). I sequenced both strands using the ABI Prism BigDye Chemistry (Perkin-Elmer) on an ABI 377 automated sequencer. I analyzed three overlapping clones each for the neo-Y and the neo-X chromosomes. I obtained the homologous region in *D. pseudoobscura* by PCR using the Expand Long PCR kit (Roche). I localized the clones to the neo-sex chromosomes of *D. miranda* by *in situ* hybridization[14]. I determined exon-intron boundaries using the homologous *D. melanogaster* release 3 sequence[15] as a reference.

Sequence variation at the neo-X copy. I used allele-specific primers for PCR amplification of the neo-X genomic regions from male genomic DNA and then carried out direct sequencing of both strands of the PCR products. I investigated sequence variability of five non-overlapping genomic fragments. Sequencing primers were spaced about every 600 bp. For the population survey, I sequenced twelve strains of *D. miranda*[3] using the ABI Prism BigDye Chemistry (Perkin-Elmer) on an ABI 377 automated sequencer.

Data analysis. I aligned sequences of *D. miranda* and *D. pseudoobscura* manually. I calculated estimates of evolutionary parameters, including θ and π (two estimators of the nucleotide diversity based on the number of segregating sites and the mean pairwise difference between alleles, respectively) and Tajima's *D* statistic for measuring departure of the site frequency spectrum from neutrality, using the program DnaSP. I calculated rates of replacement substitutions, d_N, and synonymous substitutions, d_S, for the neo-X and neo-Y branch using a maximum-likelihood method as implemented in the PAML software package, which accounts for unequal transition and transversion rates and unequal base and codon frequencies. I obtained gene functions and EST matches from Flybase.

URLs. The DnaSP program can be downloaded at http://www.ub.es/dnasp/, and the PAML software package at http://abacus.gene.ucl.ac.uk/software/paml.html. Flybase can be found at http://flybase.bio.indiana.edu/.

ACKNOWLEDGMENTS

I am grateful to P. Andolfatto, B. Charlesworth and S. Wright for discussion and comments on the manuscript. This work was supported by a European Molecular Biology Organisation postdoctoral fellowship and a Royal Society research grant.

COMPETING INTERESTS STATEMENT

The author declares that she has no competing financial interests.

Received 23 January; accepted 21 April 2003
Published online 18 May 2003; doi:10.1038/ng1164

1. Barton, N.H. & Charlesworth, B. Why sex and recombination? *Science* **281**, 1986–1990 (1998).
2. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
3. Bachtrog, D. & Charlesworth, B. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323–326 (2002).
4. Rice, W.R. Experimental tests of the adaptive significance of sexual recombination. *Nat. Rev. Genet.* **3**, 241–251 (2002).
5. Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, Cambridge, 1978).
6. Bell, G. *The masterpiece of nature* (University of California, Berkeley, 1982).
7. Charlesworth, B. & Charlesworth, D. The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000).
8. Bull, J.J. *Evolution of Sex Determining Mechanisms* (Benjamin Cummings, Menlo Park, California, 1983).
9. MacKnight, R.H. The sex-determining mechanism of *Drosophila miranda*. *Genetics* **24**, 180–201 (1939).
10. Steinemann, M. & Steinemann, S. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica* **102-103**, 409–420 (1998).
11. Rice, W.R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
12. Bachtrog, D. & Charlesworth, B. Reduced levels of microsatellite variability on the neo-Y chromosome of *Drosophila miranda*. *Curr. Biol.* **10**, 1025–1031 (2000).
13. Orr, H.A. & Kim, Y. An adaptive hypothesis for the evolution of the Y chromosome. *Genetics* **150**, 1693–1698 (1998).
14. Bachtrog, D. & Charlesworth, B. On the genomic location of the *exuperantia 1* gene in *Drosophila miranda*—the limits of *in situ* hybridisation experiments. *Genetics* (in the press).
15. Adams, M.D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
16. Li, W. *Molecular evolution* (Sinauer Associates, Sunderland, Massachusetts, 1997).
17. Bachtrog, D. Accumulation of *spock* and *worf*, two novel non-LTR retrotransposons on the neo-Y chromosome of *Drosophila miranda*. *Mol. Biol. Evol.* **20**, 173–181 (2003).
18. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge University Press, Cambridge, 1983).
19. Steinemann, M. & Steinemann, S. Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. *Proc. Natl. Acad. Sci. USA* **89**, 7591–7595 (1992).
20. McDonald, J.H. & Kreitman, M. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**, 652–654 (1991).
21. Hudson, R.R., Kreitman, M. & Aguade, M. A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**, 153–159 (1987).
22. Maynard Smith, J. & Haigh, J. The hitch-hiking effect of a favourable gene. *Genet. Res.* **23**, 23–35 (1974).
23. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
24. Rice, W.R. Degeneration of a nonrecombining chromosome. *Science* **263**, 230–232 (1994).
25. Rice, W.R. & Chippindale, A.K. Sexual recombination and the power of natural selection. *Science* **294**, 555–559 (2001).
26. Parisi, M. *et al.* Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* **299**, 697–700 (2003).
27. Carvalho, A.B. Origin and evolution of the *Drosophila* Y chromosome. *Curr. Opin. Genet. Dev.* **12**, 664–668 (2002).
28. Ohno, S. *Evolution by gene duplication* (Springer, Berlin, 1970).