Birth of a new gene on the Y chromosome of *Drosophila melanogaster*

Antonio Bernardo Carvalho[a,1], Beatriz Vicoso[a,b], Claudia A. M. Russo[a], Bonnielin Swenor[c], and Andrew G. Clark[c,1]

[a]Departamento de Genética, Universidade Federal do Rio de Janeiro, Caixa Postal 68011, 21941-971, Rio de Janeiro, Brazil; [b]Institute of Science and Technology Austria, A-3400 Klosterneuburg, Austria; and [c]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853

Contributed by Andrew G. Clark, August 21, 2015 (sent for review June 26, 2015; reviewed by Casey M. Bergman)

Contrary to the pattern seen in mammalian sex chromosomes, where most Y-linked genes have X-linked homologs, the *Drosophila* X and Y chromosomes appear to be unrelated. Most of the Y-linked genes have autosomal paralogs, so autosome-to-Y transposition must be the main source of *Drosophila* Y-linked genes. Here we show how these genes were acquired. We found a previously unidentified gene (*flagrante delicto Y*, *FDY*) that originated from a recent duplication of the autosomal gene *vig2* to the Y chromosome of *Drosophila melanogaster*. Four contiguous genes were duplicated along with *vig2*, but they became pseudogenes through the accumulation of deletions and transposable element insertions, whereas *FDY* remained functional, acquired testis-specific expression, and now accounts for ∼20% of the *vig2*-like mRNA in testis. *FDY* is absent in the closest relatives of *D. melanogaster*, and DNA sequence divergence indicates that the duplication to the Y chromosome occurred ∼2 million years ago. Thus, *FDY* provides a snapshot of the early stages of the establishment of a Y-linked gene and demonstrates how the *Drosophila* Y has been accumulating autosomal genes.

Y chromosome | *Drosophila melanogaster* | PacBio | *FDY* | *vig2*

The mammalian Y chromosome has the lowest gene density of any chromosome, and most of its genes have a homolog on the X. This pattern is consistent with the mammalian sex chromosomes having originated from an ordinary pair of chromosomes, followed by massive gene loss from the Y (1–4). In contrast, the closest homologs of all *Drosophila melanogaster* Y-linked protein-encoding genes are autosomal, strongly suggesting that its Y chromosome has been acquiring genes from the autosomes (5–7). Indeed, gene gains, and not gene losses, have played the major role in shaping the gene content of the *Drosophila* Y, at least in the last ∼63 million years (My) (8, 9). Hence, the *Drosophila* Y chromosome seems to be evolving noncanonically (10) and is an ideal model to investigate the dynamics of gene gain on a nonrecombining Y chromosome.

The *Drosophila* Y chromosome has long been known to contain genes essential for male fertility (11, 12). Due to its heterochromatic state, progress in the molecular identification of the Y-linked single-copy genes has been slow. *male fertility factor kl5* (*kl-5*), the first single-copy gene identified, was found serendipitously; it encodes a motor protein (dynein heavy chain) required for flagellar beating (13). More recently, a combination of computational and experimental methods identified 11 single-copy Y-linked genes among the unmapped sequence scaffolds produced by the *Drosophila* Genome Project (5–7). These genes have two striking features: (*i*) their closest paralogs are autosomal and not X linked, and (*ii*) they have male-specific functions, such as the beating of sperm flagella reported for the *kl-5* gene (14). The most likely explanation for this pattern is that Y-linked genes were acquired from the autosomes and have been retained because they confer a specific fitness advantage to their carriers. An autosomal origin has previously been reported for a few Y-linked genes in humans and a repetitive gene on the *Drosophila* Y (4, 15). However, unequivocal evidence of the autosomal origin of *Drosophila* Y-linked genes, and of the specific mechanism that originated them, is lacking due to their antiquity. The 11 known single-copy genes (*kl-2*, *kl-3*, *kl-5*, *ARY*, *WDY*, *PRY*, *Pp1-Y1*, *Pp1-Y2*, *Ppr-Y*, *ORY*, and *CCY*) represent ancient duplications, with amino acid identities to the putative ancestors ranging from 30% to 74%, and poor (if any) alignment at the nucleotide level. Most of them have introns in conserved positions compared with their autosomal paralogs, ruling out retrotransposition and suggesting DNA-based duplication as the mechanism. The original size of these putative duplications is unknown, because the similarity between autosomal and Y-linked regions is restricted to one gene in each case. Flanking sequences and contiguous genes either were not duplicated or were subsequently mutated and deleted beyond recognition.

Here we describe *flagrante delicto Y* (*FDY*), a single copy Y-linked gene present only in *D. melanogaster*, and which is 98% identical at the nucleotide level to the autosomal gene *vig2*. Because its origin is very recent (it occurred after the split between *D. melanogaster* and *Drosophila simulans*, ∼4 Mya), it was possible to demonstrate that *FDY* arose from a DNA-based duplication of chromosome 3R to the Y: the duplicated segment spans 11 kb of autosomal sequence and includes five contiguous genes (*vig2*, *Mocs2*, *CG42503*, *Clbn*, and *Bili*); the last four genes became pseudogenes by rapid accumulation of deletions, point mutations, and transposable element insertions or by lack of expression. Thus, *FDY* unequivocally demonstrates that the *Drosophila* Y has acquired genes from autosomes. Several Y-linked genes such as *kl-2*, *kl-3*, and *PRY* are shared by distant *Drosophila* species that diverged ∼60 Mya, implying ancient acquisitions. *FDY* dates the more recent acquisition to ∼2 My, and hence strongly suggests that *Drosophila* Y has been continuously acquiring autosomal genes.

Significance

Mammalian Y chromosomes are believed to evolve mainly through gene inactivation and loss. *Drosophila* Y chromosomes seem to not obey this rule, as gene gains are the dominating force in their evolution. Here we describe *flagrante delicto Y* (*FDY*), a very young gene that shows how Y-linked genes were acquired. *FDY* originated 2 million years ago from a duplication of a contiguous autosomal segment of 11 kb containing five genes that inserted into the Y chromosome. Four of these autosome-to-Y gene copies became inactivated ("pseudogenes"), lost part of their sequences, and most likely will disappear in the next few million years. *FDY*, originally a female-biased gene, acquired testis expression and remained functional.

Author contributions: A.B.C. designed research; A.B.C., B.V., C.A.M.R., and B.S. performed research; A.B.C. and A.G.C. analyzed data; and A.B.C. and A.G.C. wrote the paper.

Reviewers included: C.M.B., University of Manchester.

The authors declare no conflict of interest.

Data deposition: Nucleotide sequence data reported in this paper are available in DDBJ/EMBL/GenBank databases under the accession nos. TPA: BK009348–BK009355, TPA: BK009307, and KR781487.

[1]To whom correspondence may be addressed. Email: bernardo@biologia.ufrj.br or ac347@cornell.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1516543112/-/DCSupplemental.