Title: Evolutionary steps of sex chromosomes reflected in retrogenes

Authors' names: Aoife McLysaght

Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin, Ireland

Tel: +353-1-8963161

Fax: +353-1-6798558

Corresponding Author: McLysaght, A. (aoife.mclysaght@tcd.ie)

It has been shown that selective pressure to compensate for the silencing of the sex chromosomes during male meiosis resulted in many X-linked genes being duplicated as functional retrogenes on autosomes. Sex chromosome silencing in males was probably stratified during evolution, in accordance with the stratified diversification of the sex chromosomes. Here I show that the timing of the retrocopying events is associated with the timing of the X-Y differentiation of the region of X chromosome housing the parental copy of the gene.

The evolution of mammalian sex chromosomes

Mammalian sex chromosomes evolved from a pair of autosomes [1, 2]. The differentiation of the mammalian X and Y chromosomes proceeded through four (or five) steps occurring at different evolutionary times that progressively suppressed normal meiotic recombination between the X and Y, except at the pseudoautosomal regions [3-5]. Thus, the X-Y (pseudo)autosomal region was progressively shrunk by these rearrangement steps. Lahn and Page described four "evolutionary strata" on the human X chromosome created by inferred inversions on the Y, *i.e.*, four age classes of X-Y homologs [3].

The Y chromosome has degenerated substantially since it diverged from the ancestral autosome, and the modern human Y chromosome has few genes with detectable homology to those on the X chromosome [4, 5]. The imbalance in gene content between these chromosomes causes an imbalance in genes between males and females. In mammals the difference in dosage is compensated for by somatic X chromosome inactivation (XCI) in females. Sex chromosome inactivation also takes place during spermatogensis and is likely initiated by a general phenomenon called meiotic silencing of unsynapsed chromatin (MSUC), which might be involved in protecting the genome from invasion by transposable elements [6-8]. Experimental results demonstrated that MSUC silencing does not act on a chromosome as a whole, rather it acts locally on unsynapsed regions through histone modification of the DNA [7]. The MSUC phenomenon was first observed in the silencing of the sex chromosomes during male meiosis (*i.e.*, during spermatogenesis) due to the absence of recombination between the X and Y chromosomes in all but the pseudoautosomal region, where it is termed meiotic sex chromosome inactivation (MSCI) [8]. During male meiosis, the X and Y chromosomes are packaged into a so-called XY body and silenced. However,

MSUC acts locally, suggesting that the ancestral situation did not involve silencing of the sex chromosomes in their entirety.

The inactivation of the sex chromosomes during male meiosis has resulted in a dearth of genes involved in late spermatogenesis on the X chromosome and the "demasculinization" of that chromosome [9]. Recent evidence has shown increased fixation of functional retrogenes originating from the X chromosome during mammalian evolution [10-12], mirroring a phenomenon observed in *Drosophila* [13]. Kaessmann and colleagues demonstrated that more functional retrogenes were copied from parental housekeeping genes residing on the X chromosome, and that the retrocopies were biased towards expression in testis, specifically during the meiotic and post-meiotic phases of spermatogenesis [11, 12]. They showed that these events have been occurring throughout therian mammalian evolution, and dated the retrocopy events based on shared presence in different mammalian genomes. The absence of this pattern before the divergence of the marsupial lineage, as well as the unusual sex chromosomes of the platypus, indicate that the mammalian sex chromosomes are younger than previously estimated [12, 14].

Kaessmann and colleagues advocate the hypothesis that there was strong selection pressure to relocate certain genes to autosomes to compensate for the absence of expression during MSCI because they are important for spermatogenesis [15]. This hypothesis predicts that the selection pressure to move out of the X onto an autosome only appeared after the genes began to be silenced during male meiosis. In other words, genes required for spermatogenesis will have been pushed out of the X after the suppression of recombination between X and Y. The region of suppression of X-Y recombination has grown through the four events described by Lahn and Page, and therefore, the different strata of the X chromosome will have been subjected to selection pressure to retrieve testis function of genes starting at different evolutionary times. This predicts that the age of an out-of-X retrogene should correlate with the evolutionary stratum upon which the parental X chromosome gene is located.

Younger retrogenes originated from younger strata

Several studies identified functional retrogenes originating from the X chromosome and inferred the timing of the retrocopying events relative to major lineage divergences (Table 1; refs [12, 16, 17]). I related each of these parental genes to one of Lahn and Page's four evolutionary strata based on their location on the X chromosome relative to the X-linked genes used to infer the strata (Fig.1). One X chromosome gene, *RPL36A*, has given rise to four functional retrogenes located on autosomes in the human genome and I excluded this gene from the analysis because of the ambiguity it introduces, though this does not change the overall conclusions.

The data in Table 1 and Figure 1 show that of the 13 genes inferred to have retrocopied from the X on the eutherian mammal lineage before the divergence of dog, 11 of these are located on stratum 1, the oldest stratum. The three genes that retrocopied off the X between the dog and human–mouse divergence were also from stratum 1. Three of the five genes that retrocopied off the X chromosome in the human lineage after the divergence with mouse are on stratum 3, the youngest stratum where retrocopying was detected.

I tested the association between retrocopy branch (A, B, or C) and X stratum of the parental gene (*i.e.* 1, 2, 3, or 4) using Fisher's Exact Test for Count Data on the 3x4 matrix of observations (Fig. 1C). Fisher's Exact Test tests the null hypothesis of no association (independence) between counts in categorical data. It is more accurate than the Chi-Squared test or G-test when the expected numbers are small, as is the case here. The differences in the sizes of categories is implicit in the analysis and does not add bias. The association between the age of the retrocopy, and the age of the stratum housing the parental X chromosome gene is highly significant ($p << 0.001$, Fisher's Exact Test). If we reduce the complexity of the data by excluding stratum 4, which is not observed in the retrogene set, and merging stratum 2 with stratum 1 (for which initial age estimates were completely overlapping), then the association remains significant ($p<0.01$, Fisher's Exact Test). Thus, the tendency for genes to be retrocopied off the X spread stepwise through the chromosome as recombination was suppressed with the Y.

Not all retrogenes are located on the stratum "expected" under this hypothesis (or equally, they were not copied at the expected time). There is no evidence that this is due to rearrangements because the human X chromosome has a very similar arrangement to the inferred ancestral eutherian X chromosome [18], and local gene order is also well conserved across therian mammals for all of these genes. The retrocopying of these genes can be understood in the context of the low background rate observed across mammalian genomes and the ongoing nature of the selection pressure to relocate off the X [10-12].

Concluding remarks

Recent reports have shown that mammalian sex chromosomes are younger than previously estimated [12, 14]. In particular, Kaessmann and colleagues presented evidence that a bias for functional retrogenes originating from the X began in the therian mammal lineage and not before. This bias is caused by selection to retrieve late spermatogenesis functionality of X-linked genes during MSCI. The analysis presented here corroborates and extends this hypothesis by demonstrating that the pressure to relocate genes to retrieve male reproductive function did not affect the whole X chromosome simultaneously, but occurred after meiotic recombination was abolished by large chromosomal rearrangements on the Y.

Acknowledgements

I thank Henrik Kaessmann for helpful discussion and a critical appraisal of this manuscript. I also thank Mario Fares, Andrew Lloyd and Ken Wolfe for useful comments. This work is supported by Science Foundation Ireland (SFI).

Figure legend

Figure 1:

(A) Schematic representation of the human X chromosome. Genes listed on the right were retrocopied onto autosomes during mammalian evolution (refs. [12, 16, 17]). The timing of the retrocopying event (refs. [12, 16, 17]) is indicated by the colored circle beside each gene name, where the color refers to the branch color in Fig. 1B. The order and approximate location of genes are indicated, but distances are not to scale. The four evolutionary strata are shaded and numbered. The operational definitions of the four strata are indicated on the left of the figure and were inferred from the locations of genes used to define them (refs [3, 4]). The border between strata 3 and 4 is fuzzy and only the unambiguous co-ordinates are given. No retrogenes were copied from parental genes in the ambiguous regions, i.e. they were not located between genes demarcating different strata or in the region between strata 3 and 4. Pseudoautosomal regions are not indicated.

(B) Schematic phylogenetic tree of mammals. Branch lengths are not to scale, but the estimated dates of major divergences are indicated [19, 20]. Branches A, B, and C are colored red, blue and yellow, respectively.

(C) Matrix summarizing the counts of observed retrogenes in terms of the branch of the tree where the copying event occurred and the stratum on which the parental X chromosome gene is located.

Table 1: Human X-linked genes that gave rise to functional retrogenes: location and timing of retrocopying

Gene[a]	Description	Location[b]	Branch[c]	Stratum[d]	Ref.
KLHL13	Kelch-like protein 13	116.9 Mb	A	1	[12]
RRAGB	Ras-related GTP binding B	55.7 Mb	A	2	[12]
NXT2	nuclear transport factor 2-like export factor 2	108.6 Mb	A	1	[12]
ARD1A	N-terminal acetyltransferase complex ARD1 subunit homolog A	152.8 Mb	A	1	[12]
PGK1	Phosphoglycerate kinase 1	77.2 Mb	A	1	[12]
TAF7L	TAF7-like RNA polymerase II	100.4 Mb	A	1	[12]
PDHA1	pyruvate dehydrogenase (lipoamide) alpha 1	19.3 Mb	A	3	[12]
PRPS1	phosphoribosyl pyrophosphate synthetase 1	106.8 Mb	A	1	[12]
CETN2	centrin, EF-hand protein, 2	151.7 Mb	A	1	[12]
RPL10	ribosomal protein L10	153.3 Mb	A	1	[12]
TMEM185A	Transmembrane protein 185A	148.5 Mb	A	1	[12]
TAF9B	Transcription initiation factor TFIID subunit 9B	77.3 Mb	A	1	[12]
NUP62CL	nucleoporin 62kDa C-terminal like	106.3 Mb	A	1	[12]
RBMX	RNA binding motif protein, X-linked	135.8 Mb	B	1	[12, 17]
MCTS1	malignant T cell amplified sequence 1	119.6 Mb	B	1	[12]
FAM50A	Protein FAM50A	153.3 Mb	B	1	[12]
TRAPPC2	trafficking protein particle complex 2	13.6 Mb	C	3	[12]
GK	glycerol kinase	30.6 Mb	C	3	[12]
EIF2S3	eukaryotic translation initiation factor 2, subunit 3 gamma	23.9 Mb	C	3	[12]
KIF4A	kinesin family member 4A	69.4 Mb	C	1	[17]
EIF2S3	Eukaryotic translation initiation factor 2 subunit 3	23.9 Mb	C	3	[17]
TAF1	Transcription initiation factor TFIID subunit 1	70.5 Mb	C	1	[16]

[a] HGNC gene symbol
[b] Nucleotide co-ordinates from Ensembl v.49
[c] Branch of mammalian tree where retrocopying occurred. Labels as in Fig. 1B
[d] Evolutionary stratum on X chromosome.

References:

1. Ohno, S. (1967) *Sex Chromosomes and Sex-linked Genes*. Springer-Verlag
2. Graves, J.A. (1995) The evolution of mammalian sex chromosomes and the origin of sex determining genes. *Philos Trans R Soc Lond B Biol Sci* 350, 305-311; discussion 311-302
3. Lahn, B.T., and Page, D.C. (1999) Four evolutionary strata on the human X chromosome. *Science* 286, 964-967
4. Skaletsky, H.*, et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825-837
5. Ross, M.T.*, et al.* (2005) The DNA sequence of the human X chromosome. *Nature* 434, 325-337
6. Huynh, K.D., and Lee, J.T. (2005) X-chromosome inactivation: a hypothesis linking ontogeny and phylogeny. *Nat Rev Genet* 6, 410-418
7. Turner, J.M.*, et al.* (2005) Silencing of unsynapsed meiotic chromosomes in the mouse. *Nat Genet* 37, 41-47
8. Turner, J.M. (2007) Meiotic sex chromosome inactivation. *Development* 134, 1823-1831
9. Khil, P.P.*, et al.* (2004) The mouse X chromosome is enriched for sex-biased genes not subject to selection by meiotic sex chromosome inactivation. *Nat Genet* 36, 642-646
10. Emerson, J.J.*, et al.* (2004) Extensive gene traffic on the mammalian X chromosome. *Science* 303, 537-540
11. Vinckenbosch, N.*, et al.* (2006) Evolutionary fate of retroposed gene copies in the human genome. *Proc Natl Acad Sci U S A* 103, 3220-3225
12. Potrzebowski, L.*, et al.* (2008) Chromosomal Gene Movements Reflect the Recent Origin and Biology of Therian Sex Chromosomes. *PLoS Biol* 6, e80
13. Betran, E.*, et al.* (2002) Retroposed new genes out of the X in Drosophila. *Genome Res* 12, 1854-1859
14. Veyrunes, F.*, et al.* (2008) Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* 18, 965-973
15. Wang, P.J. (2004) X chromosomes, retrogenes and their role in male reproduction. *Trends Endocrinol Metab* 15, 79-83
16. Wang, P.J., and Page, D.C. (2002) Functional substitution for TAF(II)250 by a retroposed homolog that is expressed in human spermatogenesis. *Hum Mol Genet* 11, 2341-2346
17. Marques, A.C.*, et al.* (2005) Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol* 3, e357
18. Hillier, L.W.*, et al.* (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695-716
19. Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat Rev Genet* 3, 838-849
20. van Rheede, T.*, et al.* (2006) The platypus is in its place: nuclear genes and indels confirm the sister group relation of monotremes and Therians. *Mol Biol Evol* 23, 587-597