5 Veeramachaneni, V. *et al.* (2004) Mammalian overlapping genes: the comparative perspective. *Genome Res.* 14, 280–286
6 Misra, S. *et al.* (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* 3, research0083.1-0083.22
7 Yu, P. *et al.* (2005) Nested genes in the human genome. *Genomics* 86, 414–422
8 Lynch, M. and Conery, J.S. (2003) The origins of genome complexity. *Science* 302, 1401–1404
9 Lynch, M. (2006) The origins of eukaryotic gene structure. *Mol. Biol. Evol.* 23, 450–468
10 Yi, S.V. (2006) Non-adaptive evolution of genome complexity. *Bioessays* 28, 979–982
11 Lynch, M. (2007) The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8597–8604
12 Lynch, M. (2002) Intron evolution as a population-genetic process. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6118–6123
13 Lynch, M. (2007) *The Origins of Genome Architecture,* Sinauer Associates, Inc
14 Mattick, J.S. and Makunin, I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.* 14, R121–R132
15 Drosophila 12 Genomes Consortium (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218
16 Henikoff, S. and Eghtedarzadeh, M.K. (1987) Conserved arrangement of nested genes at the Drosophila Gart locus. *Genetics* 117, 711–725
17 Habib, A.A. *et al.* (1998) The OMgp gene, a second growth suppressor within the NF1 gene. *Oncogene* 16, 1525–1531
18 Jaworski, D.M. *et al.* (2007) Potential regulatory relationship between the nested gene DDC8 and its host gene tissue inhibitor of metalloproteinase-2. *Physiol. Genomics* 28, 168–178
19 Furia, M. *et al.* (1993) Dense cluster of genes is located at the ecdysone-regulated 3C puff of *Drosophila melanogaster*. *J. Mol. Biol.* 231, 531–538
20 Davies, W. *et al.* (2004) Expression patterns of the novel imprinted genes Nap1l5 and Peg13 and their non-imprinted host genes in the adult mouse brain. *Gene Expr. Patterns* 4, 741–747
21 Crampton, N. *et al.* (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res.* 34, 5416–5425
22 Osato, N. *et al.* (2007) Transcriptional interferences *in cis* natural antisense transcripts of humans and mice. *Genetics* 2, 1299–1306
23 Da Lage, J.L. *et al.* (2003) A nested alpha-amylase gene in *Drosophila ananassae*. *J. Mol. Evol.* 57, 355–362
24 Carmel, L. *et al.* (2007) Patterns of intron gain and conservation in eukaryotic genes. *BMC Evol. Biol.* 7, 192
25 ENCODE Project Consortium (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816

0168-9525/$ – see front matter
doi:10.1016/j.tig.2008.08.003 Available online 5 September 2008

Genome Analysis

Evolutionary steps of sex chromosomes are reflected in retrogenes

Aoife McLysaght

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

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. The silencing of male sex chromosomes was probably stratified during evolution, in accordance with their stratified diversification. Here I show that the timing of the retrocopying events is associated with the timing of the X-Y differentiation of the region of the 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 [3] 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).

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 probably 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) caused by 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 'demas-

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

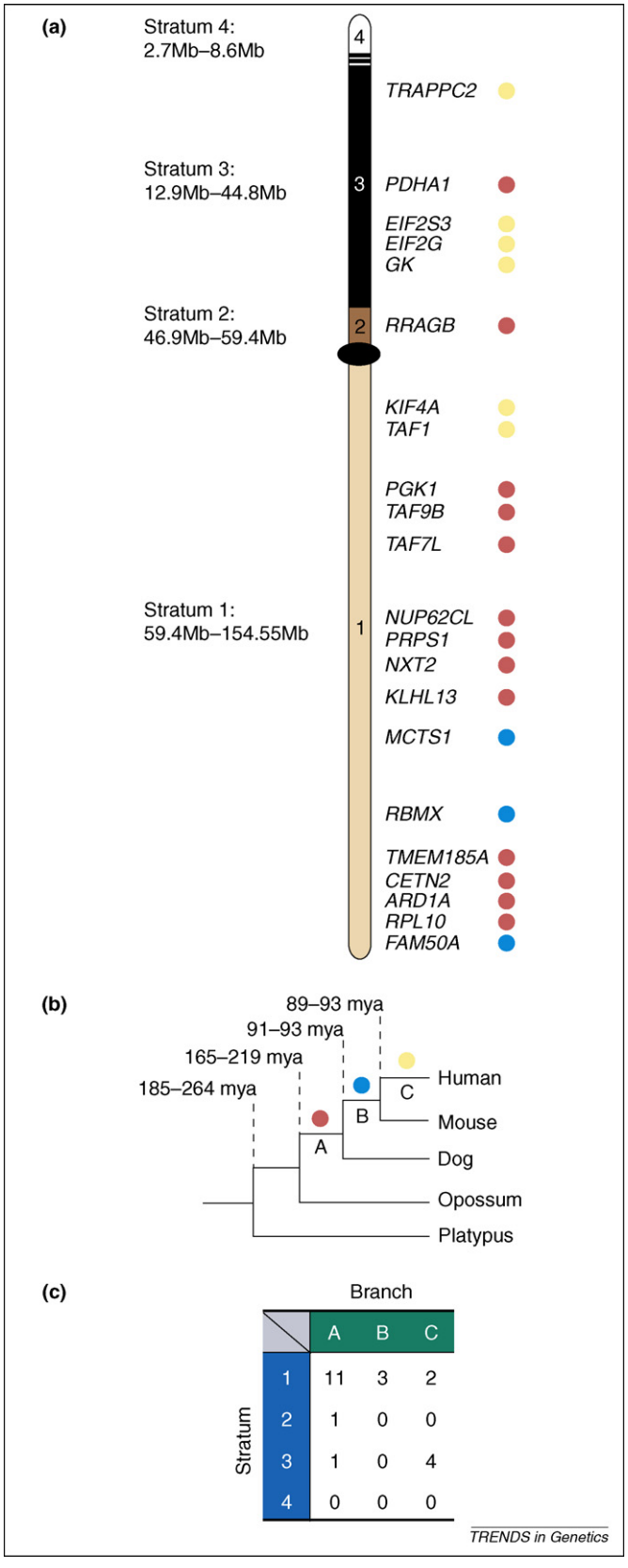

Figure 1. **(a)** Schematic representation of the human X chromosome. Genes listed on the right were retrocopied onto autosomes during mammalian evolution (Refs [12,16,17]). The timing of the retrocopying event (Refs [12,16,17]) is indicated by the colored circle beside each gene name, where the color refers to the branch color in (b). 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 coordinates 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.

culinization' 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 [11,12] demonstrated that more functional retrogenes were copied from parental housekeeping genes residing on the X chromosome and that the retrocopies were biased toward expression in testis, specifically during the meiotic and postmeiotic phases of spermatogenesis. 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 [12] advocate the hypothesis [15] 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. 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 [3], 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 on 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 the four evolutionary strata of Lahn and Page [3] based on their location on the X chromosome relative to the X-linked genes used to infer the strata (Figure 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, although this does not change the overall conclusions.

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]	Refs
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) α 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 62-kDa 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 γ	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 coordinates from Ensembl v.49.
[c]Branch of mammalian tree where retrocopying occurred. Labels as in Figure 1b.
[d]Evolutionary stratum on X chromosome.

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 the Fisher exact test for count data on the 3×4 matrix of observations (Figure 1c). The Fisher exact test tests the null hypothesis of no association (independence) between counts in categorical data. It is more accurate than the chi-square 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 ($P \ll 0.001$, Fisher 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), the association remains significant ($P < 0.01$, Fisher 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 caused by 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 [12] 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

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

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
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

0168-9525/$ – see front matter
doi:10.1016/j.tig.2008.07.006 Available online 5 September 2008

Genome Analysis

On the nature of human housekeeping genes

Jiang Zhu[1,2*], Fuhong He[1,2*], Songnian Hu[1] and Jun Yu[1]

[1] Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
[2] Graduate University of Chinese Academy of Sciences, Beijing, China

Using a collection of expressed sequence tag (EST) data, we re-evaluated the correlation of tissue specificity with genomic structure, phyletic age, evolutionary rate and promoter architecture of human genes. We found that housekeeping genes are less compact and older than tissue-specific genes, and they evolve more slowly in terms of both coding and core promoter sequences. Housekeeping genes primarily use CpG-dependent core promoters, whereas the majority of tissue-specific genes possess neither CpG-islands nor TATA-boxes in their core promoters.

Expressed sequence tag–based evidence

Housekeeping (HK) genes are ubiquitously expressed in all tissue and cell types and constitute the basal transcriptome for the maintenance of basic cellular functions. Partitioning transcriptomes into HK and tissue-specific (TS) genes and characterizing the two groups of genes in terms of their genomic structure, phyletic age, evolutionary rate and transcriptional regulation are fundamental to understand human transcriptomes. Many studies have revealed the structural [1,2], evolutionary [3,4] and promoter features of HK genes [5,6], but they were largely based on microarray data that tend to underestimate the number of human HK genes [7]. On the basis of publicly available expressed sequence tag (EST) data, we found that a large fraction (40%) of currently-annotated genes are universally expressed [7]. Here we used an EST-based estimate of expression breadth – the number of tissues in which a gene is expressed – to re-evaluate the nature of tissue specificity. We confirmed that HK genes are in general highly expressed [2] and evolve slower in coding sequence (CDS) [3] compared with TS genes. However, our analyses cast doubt on previous observations that HK genes have more compact structure [1,2] and reduced promoter conservation [6] than TS genes. In addition, we showed that HK genes are in general older than TS genes and have very distinct core promoter architecture.

Gene structure

We investigated the breath of expression for 17 288 human RefSeq loci [downloaded from NCBI (June 18, 2007 update)] across 18 human tissues (Ref. [7]; Supplementary Methods), and defined HK and TS genes as those expressed in all 18 tissues and in only 1 tissue, respectively. We observed that genes' length parameters are positively correlated with expression breadth (Figure 1a). The medians of genomic, transcript and CDS lengths are 28.8, 2.8 and 1.4 kb for HK genes and 7.2, 1.6 and 1.0 kb for TS genes, respectively (Table 1); all length parameters of HK genes are significantly longer than those of TS genes (Wilcoxon test, $P < 2.2 \times 10^{-16}$). Moreover, HK genes tend to have a greater number of exons (median of 11) than TS genes (median of 4; Wilcoxon test, $P < 2.2 \times 10^{-16}$). These observations contradict previous microarray-based results that a selection for compactness makes the lengths of intron, untranslated region (UTR) and CDS in HK genes shorter than those in non-HK genes [1,2].

The selection for compactness in HK genes was consistent with the 'selection for economy' hypothesis – highly

Corresponding author: Yu, J. (junyu@big.ac.cn).
* These authors contributed equally to this work