Evidence That the Human X Chromosome Is Enriched for Male-Specific but not Female-Specific Genes

Martin J. Lercher, Araxi O. Urrutia, and Laurence D. Hurst

Department of Biology and Biochemistry, University of Bath, Bath, United Kingdom

There is increasing evidence that X chromosomes have an unusual complement of genes, especially genes that have sex-specific expression. However, whereas in worm and fly the X chromosome has a dearth of male-specific genes, in mice genes that are uniquely expressed in spermatogonia are especially abundant on the X chromosome. Is this latter enrichment true for nongermline, male-specific genes in mammals, and is it found also for female-specific genes? Here, using SAGE data, we show (1) that tissue-specific genes tend to be more abundant on the human X chromosome, (2) that, controlling for this effect, genes expressed exclusively in prostate are enriched on the human X chromosome, and (3) that genes expressed exclusively in mammary gland and ovary are not so enriched. This we propose is consistent with Rice's model of the evolution of sexually antagonistic alleles.

Introduction

Increasing evidence suggests that X chromosomes in diverse species contain unusual complements of genes, especially sex-specific genes. In *Caenorhabditis elegans*, sperm-enriched and germline-intrinsic genes are nearly absent from the X chromosome (Reinke et al. 2000). Similarly, in Drosophila, there is a dearth of male-specific accessory gland protein genes on the X chromosome (Swanson et al. 2001). More generally, Drosophila's testes-specific genes tend to be especially abundant on autosomes, having been derived by retroposition from X-linked genes (Betran, Thornton, and Long 2002). This observation may be explained by natural selection favoring those new retrogenes that moved to autosomes and avoided the spermatogenesis X inactivation (Betran, Thornton, and Long 2002; Boutanaev et al. 2002). This is supported by the finding that clusters of testes-specific genes are described in the only known segment of the X chromosome devoid of the MSL-induced H4 acetylation (Boutanaev et al. 2002). The same may also apply in *C. elegans*, it too having an inactive X chromosome in the male germline (Fong et al. 2002; Kelly et al. 2002; Reuben and Lin 2002). Some credence is given to this hypothesis from the finding that in worm, the X chromosomes in the XX germline are silenced only in early meiosis (Kelly et al. 2002) and that ovary-expressed genes are present on the X chromosome (Reinke et al. 2000).

Is germline X chromosome inactivation (or more generally male-specific X chromosome–associated chromatin remodeling complexes [Boutanaev et al. 2002]) the sole cause of the unusual gene complement of X chromosomes? In contrast to the above, human genes whose mutants disrupt sexual development are especially common on the X chromosome (Saifi and Chandra 1999). Similarly, Wang et al. (2001), using a cDNA subtraction method, identified 25 mouse genes that appeared to be uniquely expressed in spermatogonia: three of these were Y linked and 10 were X linked. Were gene distribution

random, they argued that about an order of magnitude fewer X-linked genes would be expected.

Rice's Hypothesis

One interpretation (Hurst 2001; Wang et al. 2001) of this enrichment of spermatogonial genes on the mammalian X chromosome is that it is a consequence of the evolution of sexually antagonistic alleles (i.e., alleles that are beneficial to one sex but detrimental to the other). Rice (1984) noted that, despite the fact that an X chromosome spends only one third of its time in the male germline, a perfectly recessive allele of an X-linked gene that is favorable to the hemizygous sex (hereafter males) is much more likely to spread than an autosomal counterpart. This is because selection would act strongly on the hemizygously expressed favorable effects, whereas the deleterious effects in females would initially be masked, owing to heterozygosity in females. The autosomal counterpart would have all effects hidden and hence be likely to be lost.

If the allele is not perfectly recessive then for the autosomal case, the beneficial effects in males must counterbalance the deleterious effects in females. For the X-linked gene the beneficial effects could be relatively weak if the allele has no great fitness consequences in heterozygous females. Hence, even an allele with great negative fitness consequences when homozygous in females might spread. Consequentially, once the allele attains a significant frequency, the evolution of modifiers that force the gene to be expressed only in males is expected (Rice 1984). As most mutations are recessive, we expect an enrichment of male-specific genes on the X chromosome. Comparable logic predicts enrichment of male-benefit traits on the Y chromosome as well.

Support for the premise of Rice's model comes from the findings that the X chromosome appears to harbor a disproportionately large amount of variation in sexually selected traits (Reinhold 1998) and is, more generally, enriched for sexually antagonistic fitness variation (Gibson, Chippindale, and Rice 2002). These findings need not, however, reflect a greater abundance of genes of any given type on the X chromosome.

If Rice's hypothesis holds, we might make two predictions. First, genes expressed exclusively in other male-specific tissues will also be especially common on

the mammalian X chromosome, assuming there is no interaction with inactivation of the X chromosome (the X chromosome in murine spermatogenesis is inactivated probably by a highly conserved mechanism [Reuben and Lin 2002]). We examine this issue by looking at genes that are expressed exclusively in a somatic male-specific tissue, the prostate. Second, genes expressed in female-specific tissues need not be enriched on the X chromosome.

The latter is owing to the fact that, under Rice's model, two forces act antagonistically. Consider first a dominant allele that is beneficial to females but detrimental to males. As the X chromosome spends two thirds of its time in females, the favorable effects of the allele are evident more commonly than the deleterious effects in males, compared with the same dominant allele when autosomal. This acts as a force to increase the chances that a female-benefit /male-detriment allele might spread, were it X linked, and hence is a force leading to enrichment on the X chromosome of female-specific genes (after a modifier has suppressed the genes' expression in males). However, this force will be counterbalanced by the greater relative ease of female-advantageous/male-detrimental alleles to spread on autosomes when partially recessive, the X-linked version being relatively heavily counter selected from the outset owing to hemizygosity in males. Hence enrichment of female-specific genes on the X chromosome is not necessarily expected. We shall examine this issue by investigating the genomic location of genes expressed exclusively in human mammary gland or ovary.

Tissue Specificity and the Human X Chromosome

One important difference between the present analysis and all prior analyses is that we control for tissue specificity. We recently showed that on the average, genes on the X chromosome are expressed in fewer tissues than genes on autosomes (Lercher, Urrutia, and Hurst 2002). One might speculate that this may be the result of selection to minimize the deleterious effects of mutations in X-linked genes. This speculation aside, if X-linked genes do tend to be tissue specific per se, then we expect enrichment on the X chromosome for any class of genes that are tissue specific regardless of sex specificity. This could indeed go some way to explain prior results. Hence, we establish a data set of expression patterns for over 8,000 genes but then extract only those expressed in just one tissue.

Materials and Methods
The SAGE Data Set

We used publicly available data from Serial Analysis of Gene Expression (Velculescu et al. 1995; SAGE). From SAGEmap (Lash et al. 2000) at NCBI, we obtained a reliable mapping of UniGene (Schuler et al. 1996) groups to *Nla*III SAGE tags. Each UniGene group consists of all GenBank sequences representing the same human gene. In the remainder, we will refer to each such group as a gene and represent it by its longest RefSeq sequence. Tags mapping to more than one gene were excluded. We located 11,612 RefSeq genes on the August 2001 Golden Path assembly of the human genome (http://genome.

cse.ucsc.edu/), each labeled unambiguously by at least one SAGE tag. This set of gene/tag combinations was crosslinked to the quantitative expression profiles at SAGEmap. Positive expression was seen in 8,367 genes in at least one out of 35 libraries representing 14 normal (i.e., non-pathological) tissues. If a tag had been counted only once in one tissue, this was most likely due to a sequencing error, and we discounted the observation. Adding all counts for libraries representing the same tissue type, we then calculated breadth of expression (number of tissues with positive expression) for each gene. Genes were counted as tissue specific if they were expressed in only one of the 14 tissues.

Statistics

To determine the significance of the observed number of genes of a given class (prostate, ovary/mammary) on the X chromosome against null expectations, we employed a randomization strategy. We reassigned all genes at random to chromosomes while maintaining the total gene count, the total count of genes within each class, and the total number of genes on each chromosome as found in the original data set. The P value was then specified as the proportion of randomizations in which the actual number, or a greater number, of genes within the class in question appeared on the X chromosome.

The expectations for the number of genes on the X chromosome can be derived by this method or by partitioning the data into tissue-specific genes that are not sex specific and using the X:A ratio to deduce the expected number of X-linked genes within any given class, given the total number of genes in this class. Both method estimates are provided. The first estimate given below is always from the X:A ratio, and the second is from randomization.

Results

Our prior work suggested that genes on the X chromosome are not expressed in as many tissues as autosomal genes (Lercher, Urrutia, and Hurst 2002). Does it follow that the X chromosome has more tissue-specific genes? If we examine genes expressed in at least nine of the 14 tissues (N = 1,897) (our prior definition of housekeeping genes [Lercher, Urrutia and Hurst 2002]), we find 50 that are X linked (i.e., 2.7% of the total). By contrast, of genes expressed in three or fewer tissues (N = 3,441), 3.8% are X linked ($P < 0.02$ by randomization, two tailed). Of those expressed in just one tissue, 3.6% of the total of 1,511 are X linked. Although this latter result is not significant at the 5% level ($P = 0.069$, by randomization, two tailed), given the apparent tendency, it is best to be conservative and to control for tissue specificity.

Are prostate-specific genes especially prevalent on the X chromosome? Of the tissue-specific genes that are not expressed in the sex-specific tissues (ovary, mammary gland, or prostate)1,046 are autosomal and 35 (3.3%) are X linked. Of the prostate-specific genes, 189 are autosomal compared with 13 (6.9%) that are X linked. This represents an approximate doubling of the frequency of

prostate-specific genes on the X chromosome and represents a significant enrichment (6.5/ 7.3 are expected, $P = 0.02$, one tailed, derived by 100,000 randomizations). Pairwise Blast of all of the X-linked prostate-specific genes against all the others on the X chromosome revealed no duplicate genes, so the enrichment is not owing to higher rates of duplication on the X chromosome.

It may be notable that our estimate of the extent of the enrichment of male-specific genes (an approximate doubling) is lower than that observed by Wang et al. (2001). This is unlikely to be owing solely to methodological differences (of which control for tissue specificity would be one), as the difference appears to be quite large: Wang et al. report that nearly 40% of the spermatogonia-specific genes are X linked, which compares with just 7% for prostate. Perhaps there is significant heterogeneity between male-specific tissues? When high-quality expression data is available for more male-specific tissues, this should be testable.

In our sample, female-specific genes, in contrast to the male-specific genes, show no X-linked enrichment when compared against tissue-specific genes. Whereas 222 genes expressed in ovary or mammary gland are autosomal, only six (2.7%) are X-linked genes expressed in either tissue. If anything then, female-specific genes are underrepresented on the X chromosome, although the difference is not statistically significant (six observed, 7.4/8.2 are expected, $P = 0.33$). Analyzing ovary alone (under the supposition that some mammary gland genes might also be in male breast tissue) does not alter the conclusions: 107 are autosomal, four are X linked, and four are expected (by both methods) ($P = 0.57$).

Discussion

The above results provide support, by no means definitive, that Rice's hypothesis may be important to understanding mammalian X chromosome evolution. However, this should be regarded as a provisional interpretation, as numerous caveats must be noted. For example, in several years time SAGE data will, no doubt, be available for many more tissues, in which case, it is all but inevitable that some of our ''tissue-specific'' genes will turn out not to be tissue specific at all, just expressed in relatively few tissues. This need not prove be too problematic for the current provisional interpretation, as Rice's model does not require the genes to be expressed exclusively in one tissue. However, more problematically, it may yet prove to be the case that some ''ovary-specific'' genes are in fact germline-specific genes and expressed in both males and females. Prior evidence suggests that genes expressed in both germlines are not enriched on the X chromosome (Wang et al. 2001). SAGE analysis on testicular tissue would allow us to eliminate this possibility.

Further, in our presentation of Rice's hypothesis, we assumed the presence of alleles expressed in both sexes for genes already present on the X chromosome. It is uncertain whether it is reasonable to suppose that there were genes expressed both in prostate and in females as well. Similarly, it may possibly be that the genes were originally autosomal and their sexually antagonistic phenotype predisposed them to becoming X linked (Charlesworth and Charlesworth 1980). Even were our finding statistically robust, the interpretation is by no means certain.

Despite the above caveats, given the present results and those of Wang et al. (2001), we can tentatively suggests that, consistent with Rice's hypothesis, the mammalian X chromosome is enriched for male-specific but not female-specific genes. What also of the Y chromosome? As expected, in our sample, no mammary-specific or ovary-specific genes are Y linked. Two of the seven Y-linked sequences in our sample were prostate specific, the others being expressed (apparently in a sex-specific manner) either in brain or in peritoneum. Overall enrichment of prostate-specific genes on the X or Y chromosome is significant ($P = 0.01$, by randomization).

The description of some brain-specific, Y-linked genes is especially notable, as it has also recently been suggested that selection for sex differences in cognitive ability might explain why genes that affect cognitive ability appear also to be enriched on the X chromosome (Zechner et al. 2001). Although there are too few brain-specific, Y-linked genes to perform meaningful statistics, there may be weak enrichment of these: we expect about one and observe three. This and the putative X chromosome enrichment may also reflect the processes envisaged by Rice. However, brain-specific genes (white matter, astrocyte, and thalamus) in our sample are not enriched on the X chromosome: we expect 13.5/12.9 X-linked genes, which compares with 14 observed ($P = 0.43$) (of 406 brain-specific genes, 389 are autosomal and 14 [2.1%] are X linked; of non–brain-specific, non–sex-specific genes, 657 are autosomal and 21 [3.2%] are X linked). This brain sample presumably includes both sex-specific and non–sex-specific genes, and it would be valuable to return to the issue using direct expression assays when sex specificity of gene expression in non–sex-specific tissues can be assayed.

Acknowledgments

Literature Cited

Betran, E., K. Thornton, and M. Long. 2002. Retroposed new genes out of the X in Drosophila. Genome Res. **12**:1854–1859.

Boutanaev, A. M., A. I. Kalmykova, Y. Y. Shevelyou, and D. I. Nurminsky. 2002. Large clusters of co-expressed genes in the Drosophila genome. Nature **420**:666–669.

Charlesworth, D., and B. Charlesworth. 1980. Sex-differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet. Res. **35**:205–214.

Fong, Y. Y., L. Bender, W. C. Wang, and S. Strome. 2002. Regulation of the different chromatin states of autosomes and

X chromosomes in the germ line of *C. elegans*. Science **296**:2235–2238.

Gibson, J. R., A. K. Chippindale, and W. R. Rice. 2002. The X chromosome is a hot spot for sexually antagonistic fitness variation. Proc. R. Soc. Lond. B Biol. Sci. **269**:499–505.

Hurst, L. D. 2001. Evolutionary genomics—sex and the X. Nature **411**:149–150.

Kelly, W. G., C. E. Schaner, A. F. Dernburg, M. H. Lee, S. K. Kim, A. M. Villeneuve, and V. Reinke. 2002. X-chromosome silencing in the germline of *C. elegans*. Development **129**:479–492.

Lash, A. E., C. M. Tolstoshev, L. Wagner, G. D. Schuler, R. L. Strausberg, G. J. Riggins, and S. F. Altschul. 2000. SAGE-map: a public gene expression resource. Genome Res. **10**:1051–1060.

Lercher, M. J., A. O. Urrutia, and L. D. Hurst. 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat. Genet. **31**:180–183.

Reinhold, K. 1998. Sex linkage among genes controlling sexually selected traits. Behav. Ecol. Sociobiol. **44**:1–7.

Reinke, V., H. E. Smith, J. Nance et al. (11 co-authors). 2000. A global profile of germline gene expression in *C. elegans*. Mol. Cell. **6**:605–616.

Reuben, M., and R. Lin. 2002. Germline X chromosomes exhibit contrasting patterns of histone H3 methylation in *Caenorhabditis elegans*. Dev. Biol. **245**:71–82.

Rice, W. R. 1984. Sex-chromosomes and the evolution of sexual dimorphism. Evolution **38**:735–742.

Saifi, G. M., and H. S. Chandra. 1999. An apparent excess of sex- and reproduction-related genes on the human X chromosome. Proc. R. Soc. Lond. B Biol. Sci. **266**:203–209.

Schuler, G. D., M. S. Boguski, E. A. Stewart et al. (101 co-authors). 1996. A gene map of the human genome. Science **274**:540–546.

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in Drosophila. Proc. Natl. Acad. Sci. USA **98**:7375–7379.

Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. Science **270**:484–487.

Wang, P. J., J. R. McCarrey, F. Yang, and D. C. Page. 2001. An abundance of X-linked genes expressed in spermatogonia. Nat. Genet. **27**:422–426.

Zechner, U., M. Wilda, H. Kehrer-Sawatzki, W. Vogel, R. Fundele, and H. Hameister. 2001. A high density of X-linked genes for general cognitive ability: a run-away process shaping human evolution? Trends Genet. **17**:697–701.

William Jeffery, Associate Editor