

region, where two genes (*SYBL1* and *H-SPRY-3*) that are subject to X inactivation have Y-linked homologs that are also silenced (Maria-Rosaria Matarazzo, International Inst. of Genetics and Biophysics, Italy). How such Y-linked silencing is achieved remains to be discovered.

The imprinted inactivation of the paternal X chromosome² that occurs in marsupials and in the extra-embryonic tissues of rodents is often cited as the ancestral form of X inactivation. Nobuo Tagaki (Hokkaido University, Japan) presented data challenging this concept. In mouse XX androgenotes, which have two paternal X chromosomes, rather than inactivation of both Xs, random X inactivation of one or the other X occurred in extra-embryonic cells. This suggests that resistance of the maternal X to inactivation, rather than obligatory paternal X inactivation, underlies imprinted X inactivation. It also implies that a counting mechanism, capable of recognizing the number of X chromosomes in the cell, functions in extra-embryonic tissues, as in random X inactivation. However, the paternal X might still show some predisposition to inactivation because it is coated by *Xist* RNA well before the first signs of inactivation in extra-embryonic tissues. What is the nature of the X-inactivation imprint? Methylation analysis – a ‘must’ for any self-respecting imprinting story – of the *Xist* 5' region revealed no obviously imprinted sites (Graham Kay, the Queensland Institute of Medical Research, Australia), but it might be that the 3' end is where it all happens³. The CpG-rich region lying 15 kb 3' to *Xist* and close to the promoter of the recently described *Xist* antisense transcript, is one good candidate. Deletion of this region abolishes *Xist* sense and antisense expression in undifferentiated embryonic stem (ES) cells and also leads to non-random inactivation of the targeted allele (Phil Avner, Pasteur Institute, France).

X inactivation during spermatogenesis seems to be yet another story. The transient nature of this X inactivation in male meiotic cells could be owing to the absence of methylation and histone acetylation or, more simply, to transient exclusion of RNA polymerase II from the sex vesicle (XY body). Because this X inactivation is unlikely to be regulated by *Xist*, as it is apparently unaffected in *Xist* knockout mice (John McCarrey, the Southwest Foundation for Biomedical Research, USA), the significance of the very low levels of *Xist* expression and possible *Xist*-coating of the sex vesicle observed in male meiotic cells remains unresolved, as does indeed the evolutionary significance of X inactivation during gametogenesis.

Another intriguing issue concerns which parts of the *Xist* gene are involved in its regulation and function. Insights are expected to come from gene targeting and evolutionary studies. One surprise is that deletion of one of the most highly conserved regions of the *Xist* transcript, exon 4, gives no apparent phenotype (Marie-Laure Caparros, MRC, London, UK). Another surprise is that the upstream ES-cell-specific promoter P0 region is highly repetitive and poorly conserved in vole species, unlike the P1 and P2 somatic promoters (Tatyana Nesterova, Academy of Sciences, Novosibirsk, Russia). There might even be species differences in the way *Xist* is regulated during development, because human *XIST* RNA is stable and coats autosomes *in cis* in undifferentiated mouse ES and embryonal carcinoma cells, unlike its mouse counterpart (Edith Heard, Pasteur Institute, France; Ikuya Yoshida, Hokkaido University, Japan). In fact, *Xist* still hasn't passed the ultimate evolutionary test for mammals: is it conserved in marsupials?

Future directions in the X-inactivation field include a promising mutagenesis program reported by Laura Carrel, aimed at isolating additional X inactivation variants. Some of these mutations might be associated with the *Xce* locus, which has long been a paradigm for bias in X inactivation⁴ and whose map position has recently been greatly refined (Phil Avner, Pasteur Institute, France). Identification of some of the protein partners of the *Xist* transcript and the inactive X will also be crucial. To date, histone macroH2A is the only candidate. As it is apparently not required for maintenance of the inactive state and Jacqueline Mermoud (MRC, London, UK) showed that its colocalization with the inactive X occurs surprisingly late during *in vitro* differentiation, its role remains enigmatic. Integrating all this and much more suggests that understanding X inactivation, while indeed a subject fit for mammals, could require a touch of divine insight.

Further reading

- 1 Lyon, M.F. (1998) X-Chromosome inactivation: a repeat hypothesis. *Cytogenet. Cell Genet.* 80, 133–137
- 2 Latham, K.E. (1996) X-chromosome imprinting and inactivation in the early mammalian embryo. *Trends Genet.* 12, 134–138
- 3 Heard, E. *et al.* (1999) Anti-*Xist*entialism. *Nat. Genet.* 21, 343–344
- 4 Heard, E. *et al.* (1997) X-chromosome inactivation in mammals. *Ann. Rev. Genet.* 31, 571–610

Organelle genes – do they jump or are they pushed?

Race *et al.* presented a stimulating account of the likely reason for retention of genes in some organelles in September's issue¹. As they made clear, this begs the question of what drives the loss of genes to the nucleus anyway. They described suggestions that the selection pres-

sure might arise from Muller's ratchet, exacerbated by the occurrence of DNA-damaging processes in the organelles. We would like to suggest another explanation (that might act together with Muller's ratchet), based on our studies of plastid genomes. These genomes have a tendency to



Christopher J. Howe
c.j.howe@bioc.cam.ac.uk

Adrian C. Barbrook
acb18@mole.bio.cam.ac.uk

Peter J. Lockhart*
p.j.lockhart@massey.ac.nz

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, UK CB2 1QW.

*Institute of Molecular Biosciences, Massey University, Palmerston North, New Zealand.

become AT rich, both in coding and in non-coding regions². The reason for this is unknown, but it might be a consequence of the nature of the DNA damage to which the genomes are exposed. The high AT content affects the amino acid composition of the encoded proteins³. Plastid genes are enriched in codons for Phe, Ile, Lys, Asn and Tyr and depleted in codons for Ala, Gly and Pro by comparison with genes for the same proteins from the nucleus or prokaryotes³. For some proteins, this shift in their amino acid composition might be sufficiently deleterious to their function to provide a significant selective advantage for their transposition to the nucleus. Mitochondrial genomes are also AT rich⁴, so the same processes might operate there. Therefore, transposed genes can be seen as 'molecular refugees' fleeing from an oppressive genomic environment that is trying to force them into encoding unsuitable proteins. Their freedom of movement might be

limited by other factors, perhaps including membrane integrity, allowing some very AT-rich genomes (such as the mitochondrial genome of *Reclinomonas americana*) to retain a larger genetic complement than usual⁵.

References

- 1 Race, H.L. *et al.* (1999) Why have organelles retained genomes? *Trends Genet.* 15, 364–370
- 2 Lockhart, P.J. *et al.* (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol.* 34, 153–162
- 3 Barbrook, A.C. *et al.* (1998) Phylogenetic analysis of plastid origins based on *secA* sequences. *Curr. Genet.* 34, 336–341
- 4 Gray, M.W. *et al.* (1998) Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* 26, 865–878
- 5 Lang, B.F. *et al.* (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387, 493–497



Identifying HOX paralog groups by the PBX-binding region

The *Hox* genes are a family of transcription factors that define specific anteroposterior identities, both in vertebrate and in invertebrate embryos¹, and that are characterized by a very highly conserved DNA-binding motif known as the homeodomain^{2–4}. *In vitro*, most HOX proteins recognize the same four-base-pair consensus sequence that is actually repeated many times in the genome⁵. Far greater binding specificity is achieved when HOX proteins bind as heterodimers with PBX proteins (vertebrate homologs of

Drosophila homeodomain-containing transcription factor *extradenticle*)⁶. PBX and HOX proteins interact at a specific and highly conserved hexapeptide on the surface of the HOX protein^{6–8}. This short sequence of amino acids is necessary for PBX binding and, apart from the homeodomain itself, is the most characteristic feature of *Hox* genes.

During the evolution of vertebrates, the ancestral cluster of *Hox* genes was duplicated at least twice^{9–11}; hence, most vertebrates have at least four independent *Hox* clusters, referred to as a, b, c...etc. Despite some *Hox* genes in each cluster having become non-functional or even entirely deleted subsequent to the duplication step, the overall genomic structure of each cluster has been conserved in evolution¹. In general, the descendants of each of the genes in the ancestral *Hox* cluster have similar expression patterns and some conserved functions^{12,13}. They are described as paralogs (e.g. *hoxa1*, *hoxb1* and *hoxd1*). Outside of the homeodomain region, the overall sequence identity between members of each paralog group is very low. As a result, the paralog identity of each gene has often been ambiguous.

Two previous studies have addressed this problem by aligning *Hox* genes based on their hexapeptide sequences rather than their homeodomain^{14,15} (which forms the usual basis for *Hox* gene alignment comparisons). Their findings revealed that there were, indeed, some amino acids adjacent to the hexapeptide that are conserved only within individual paralog groups. Here we have extended these studies to include all hexapeptide-containing paralog groups from a wide range of species. Interestingly, this reveals that there are several very highly conserved amino acids clustered around the hexapeptide sequence. These amino acids consistently identify *Hox* genes as belonging to a particular paralog group (Fig. 1).

Why are the amino acids around the hexapeptide sequence so highly conserved between paralogs but not clusters? Members of one paralog bind to a distinct DNA sequence only when bound to PBX at the hexapeptide site;

FIGURE 1. Conserved amino acids around the hexapeptide sequence of *Hox* paralog groups (1–8)

HOX1	T F D W M K V K R N x P K
HOX2	P E F P W M K E K K x x x K
HOX3	K Q I F P W M K E S x Q x x K x K
HOX4	V V Y P W M K K h H h x x x x x Y
HOX5	P Q I Y P W M r K L H h x H x x x x x K R
HOX6	h Y P W M Q R M N S x x x x f G x x R x R
HOX7	R I Y P W M R S x G x D
HOX8	h F P W M

trends in Genetics

Capital letters denote the conserved amino acids (based on the standard IUB codes). The amino acids in bold are those that define the hexapeptide itself. 'f' denotes either phenylalanine (F) or tyrosine (Y) and 'r' denotes either arginine (R) or lysine (K). 'h' is a hydrophobic amino acid. 'x' indicates that the amino acid at that position is not conserved (although the spacing is). The sequences compared for each paralog group were as follows. HOX1: *Hoxa1* (R, H, M, X, Z), *Hoxb1* (F, H, M, X, Z) and *Hoxd1* (M, X). HOX2: *Proboscipedia* (D), *Hoxa2* (M) and *Hoxb2* (H). HOX3: *Hoxa3* (M, Z), *Hoxb3* (CP, C, H, M, X) and *Hoxd3* (C, H, M). HOX4: *Hoxa4* (C, H, M), *Hoxb4* (F, M, X, Z), *Hoxc4* (H, M, Z) and *Hoxd4* (C, H, M, Z). HOX5: *Sex combs reduced* (D), *Hoxa5* (H, M, Z), *Hoxb5* (C, H, M, X, Z) and *Hoxc5* (H, M). HOX6: *Hoxb6* (H, M, Z) and *Hoxc6* (F, H, M, X, Z). HOX7: *Hoxa7* (H, M, Z) and *Hoxb7* (H, M, X). HOX8: *Hoxb8* (C, M, X) and *Hoxc8* (F, H, M, X). Abbreviated species names: D, *Drosophila*; C, chicken; CP, carp; F, *Fugu*; H, human; M, mouse; R, rat; X, *Xenopus*; Z, Zebrafish.

Richard Morgan
morgan@ucl.ac.uk

Paul In der Rieden
rieder@niob.knaw.nl

Michiel H.W. Hooiveld
hooiveld@niob.knaw.nl

Antony J. Durston
tony@niob.knaw.nl

The Hubrecht Laboratory,
Uppsalalaan 8, 3584 CT
Utrecht, The Netherlands.