STUDIES OF X-CHROMOSOME INACTIVATION AND THE IDENTIFICATION OF THE *XIST* GENE IN THE INSECTIVORE *SCAPANUS ORARIUS*

by

SANJA KARALIĆ

B.Sc. (Hon.) Simon Fraser University, 1998

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF MEDICAL GENETICS

FACULTY OF MEDICINE

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

May 2001

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Medical Genetics

The University of British Columbia
Vancouver, Canada

Date May 30th, 2001

DE-6 (2/88)

Abstract

X-chromosome inactivation is the transcriptional silencing of one of the two X chromosomes in the mammalian female and is thought to be a means of dosage compensation. The inactive X chromosome is condensed, late replicating, hypoacetylated and hypermethylated at CpG islands. The silenced chromosome also expresses *Xist* – a fascinating large, untranslated RNA involved in the inactivation process. Although the mechanism of *Xist* function remains unknown, it has been established that the RNA is necessary for X-chromosome inactivation in two eutherian mammals – humans and mice.

Xist fragments have also been identified for several other eutherians, but the only complete sequences are those of the human, murine and vole genes. Interestingly, *Xist* has not yet been identified in marsupials, although these mammals undergo X-silencing as well. Imprinted inactivation and the lack of CpG methylation are two additional aspects of X-inactivation in which the marsupials differ from the eutherians. X-chromosome inactivation studies in the eutherian mammals have mostly been limited to the human and murine model organisms and little is known about the process in other members of this mammalian subclass.

This study examined the process of X-inactivation and identified the *Xist* gene in a virtually unstudied mammalian organism – the Coast mole, *Scapanus orarius*. The mole belongs to the order Insectivora, an ancient and possibly basal mammalian group, which was traditionally thought to be closest to the eutherian common ancestor. Prior to this study, X-chromosome inactivation in the insectivores had not been investigated in detail and the *Xist* gene had not been identified. By using methylation-sensitive restriction digests of several X-linked genes (*ARA*, *FMR1* and *ZFX*) followed by PCR with primers flanking the restriction

sites, this study confirmed the presence of a methylated X chromosome, indicating that these animals undergo X-inactivation. Additionally, the *ARA* and *FMR1* genes were found to be subject to silencing, while *ZFX* was shown to escape X-inactivation in the mole. The results of the methylation assay not only show that X-inactivation exists in the Coast mole, but that silencing through methylation is conserved within the Eutheria.

Low stringency PCR with primers designed from sequence conserved among the available *Xist* genes was successful in amplifying a mole fragment. Sequence analysis and female-specific expression confirmed that this was the Coast mole *Xist* homologue. Additional *Scapanus Xist* fragments were obtained through long range PCR and inverse PCR resulting in a total of 2.5 kilobases of sequence, which show significant similarity to *Xist* fragments from other mammals. The presence of *Xist* in the mole, along with its exclusively female expression pattern and sequence similarity to the gene in other species, indicate that *Xist* is involved in the X-inactivation process in this insectivore.

In conclusion, the study of X-chromosome inactivation in the Coast mole shows that the unique features of eutherian X-linked silencing have been conserved in the Order Insectivora.

Table of contents

List of tables

List of figures

Acknowledgements

I would like to thank my supervisor, Dr. Carolyn Brown, for her excellent guidance and support throughout my Master's programme. I would also like to thank everyone in the Brown Lab, my committee members (Dr. Wendy Robinson, Dr. Sally Otto, Dr. Marco Marra, Dr. Fred Dill, Dr. Dixie Mager and Dr. Kelly McNagny), my "mole connections", Dr. Kevin Campbell and Tim Sheehan, Dr. Jim Rupert and Jane Lee.

Chapter 1

An introduction to sex determination and sex chromosomes

X-chromosome inactivation is the transcriptional silencing of one of the two X chromosomes (XX) in the mammalian female in order to achieve dosage equivalency with the male and is thought to be a direct consequence of the mammalian system of chromosomal sex determination (Lyon, 1961). The choice of sex in the animal world takes many forms and can involve a range of chromosomal and environmental factors. Sex determination in invertebrates is varied and a complete discussion is beyond the scope of this manuscript. However, two invertebrates using the chromosomal system of sex determination, *Drosophila melanogaster* and *Caenorhabditis elegans*, are especially relevant to the understanding of dosage compensation and will be discussed later.

Although the most familiar type of sex determination in vertebrates is the XX-XY system, characteristic of most mammals, in reality the subphylum Vertebrata shows a great diversity when it comes to defining sex (for a review see Solari, 1993). Fish, the most basal and ancient of all vertebrates, show a large plasticity regarding sex determination - most species have undifferentiated sex chromosomes, while others use the XX-XY, ZW-ZZ, XX-XO, or multiple gonosome (sex chromosome) systems. These gonosome combinations are found scattered among the various species, seemingly without any strong phylogenetic connections, implying that sex chromosomes arose independently several times in the Pisces. Additionally, fish are unique among vertebrates in the number of species that are naturally hermaphroditic, either simultaneously or sequentially. Furthermore, in some gonochoristic species (separate sexes) sex

change can be easily elicited by simple hormonal treatment, while in hermaphrodites even behavioural changes are sufficient for sex reversal.

The most ancient group of extant reptiles, the turtles and crocodilians, show environmental sex determination in which the crucial switch seems to be temperature. Heterogamety seems to have arisen several times in reptiles and species exhibit simple male heteromorphism (XX-XY), simple female heteromorphism (ZW-ZZ) or depend on multiple sex chromosomes. The degree of differentiation of the sex chromosomes varies widely and ranges from the cytogenetically indistinguishable to the visibly different, as is the situation in most snakes. While the knowledge of reptilian linkage maps remains scarce, it does not seem that the X or Z chromosomes have been conserved among the various reptilian groups, possibly due to the lack of dosage compensation in the form of X inactivation (Solari, 1993).

Compared to other vertebrates, the situation in the Classes Mammalia and Aves is relatively straightforward. Both groups possess well-differentiated heteromorphic sex chromosomes, which seem to have arisen from an ancestral autosomal pair (Pask and Graves, 1999). In mammals the male is the heterogametic sex (XY *vs.* XX), while in birds it is the female who carries two types of sex chromosomes (ZW *vs.* ZZ). The X and Z chromosomes are relatively large and conserved, while the Y and W seem to show a range of differentiation (Graves et al., 1998a). In eutherian and marsupial mammals and carinate (flying) birds, the Y and W are small, heterochromatic and may pair over a small region with the X or Z. On the other hand, the Y in the monotremes and W in the ratite (flightless) birds are large and may pair completely with their respective partner (Graves et al., 1998a).

However, significant differences exist between these two systems - the most glaring discrepancy is the identity of the heterogametic sex. In most mammals sex is controlled by a

dominant testis determining factor, identified as *SRY* (sex determining region on the Y) by Sinclair et al. in 1990, while avian sex seems to depend on a gene-dosage effect of the Z-linked *DMRT1* (doublesex and mab-3 related transcription factor) gene (Nanda et al., 2000). A recent study in turtles has shown that *DMRT1* is the first gene to show temperature-dependent expression prior to sexual differentiation in reptiles (Kettlewell et al., 2000). Interestingly, *DMRT1* is related to genes involved in sex regulation in both *Drosophila* and *C. elegans* and the mammalian *DMRT1* homologue is located on the sex-reversal region of chromosome 9 (Nanda et al. 2000). These observations have led Nanda and colleagues (2000) to speculate that *DMRT1* may be the dosage-based sex determination gene in ancestral vertebrates. Furthermore, while mammals engage in X-chromosome inactivation to achieve dosage equivalency of essential genes between males and females, birds do not seem to compensate for the gene imbalances arising from the possession of heteromorphic sex chromosomes (Baverstock, 1982). Comparative mapping of the avian Z chromosome and mammalian X chromosome has shown that most bird Z-linked genes map to mammalian autosomes, especially human chromosome 9 (Nanda et al. 2000), and *vice versa*, indicating that these two sex chromosome systems evolved independently from different autosomal pairs (Ohno, 1967, Graves et al., 1998a). Unfortunately, it seems impossible to deduce the sex determining mechanism of a common reptilian ancestor that would have existed approximately 350 million years ago (Pask and Graves, 1999), since modern reptiles exhibit a wide range of chromosomal and environmental sex determination.

As is evident from this overview of vertebrate sex determination, sex chromosomes have arisen many times in many different phyla. It is now widely believed that the vertebrate proto-gonosomes were a pair of autosomes in which one allele acquired a sex-determining function,

followed by a suppression of recombination and gradual degradation of most of the chromosome carrying the novel gene, eventually leading to a fully differentiated pair of sex chromosomes. Evidence of this process can be seen in the residual identities between the sex chromosomes and by the identification of species that seem to be at different stages in this differentiation process from a homomorphic to a heteromorphic chromosome pair, such as snakes or ratite and carinate birds (Solari, 1993). The problem of dosage compensation that may arise in organisms with heteromorphic sex chromosomes has been dealt with in three manners represented by the nematode *Caenorhabditis elegans*, the fly *Drosophila melanogaster* and the eutherian mammal, such as mouse or humans.

In *C. elegans* the XX genotype defines a functional hermaphrodite, while an XO animal is a male. Gene product equivalency has been achieved by down regulating gene expression through a multi-protein complex that assembles on both X chromosomes of the hermaphrodite, presumably reducing the transcriptional activity by altering chromatin structure (Lucchesi, 1998, Meller, 2000). The mechanism by which this dosage compensation machinery recognizes the correct chromosome is unknown, but may have something to do with the organization of the *C. elegans* chromosomes. Unlike the autosomes, the X chromosome lacks a central gene cluster, and one type of repetitive element is less frequent and more uniformly distributed on the X in comparison to the autosomes (Meller, 2000).

The fruit fly deals with the problem of dosage compensation in the opposite manner by up-regulating expression from the XY male's single X chromosome, while expression from both of the female's X's remains unchanged. This increase in transcription is achieved through the MSL (male specific lethal) complex, composed of five identified proteins and two non-coding

roX RNAs (*roX1* and *roX2*), which appears to remodel chromatin by site-specific acetylations of histone H4 (Kelley and Kuroda, 2000a).

Mammals employ the third known alternative for dosage equivalency, which is to completely silence one of the female's two X chromosomes. The inactivated chromosome is condensed, hypoacetylated, hypermethylated, nuclease-insensitive and replicates late in the cell cycle (Heard et al., 1997). In some species the condensed chromatin is visible as a Barr body. The process of X inactivation is thought to be responsible for the widespread conservation of synteny and size of the mammalian X chromosome, especially among eutherian mammals, since fragmentation of the X would disrupt dosage compensation. This phenomenon was first observed by Susumu Ohno in 1967 and is now known as Ohno's Law. Interestingly, dosage compensation has not been observed in a large number of organisms with heteromorphic sex chromosomes, such as birds or reptiles (Solari, 1993). Another paradox is the observation that the content of the bird Z chromosome seems to be well conserved despite the lack of chromosome wide inactivation (Nanda et al., 2000).

Mammalian Sex Chromosomes

The Order Mammalia consists of two subclasses: the Prototheria (monotremes) and Theria, which is in turn divided into the Metatheria (marsupials) and Eutheria (incorrectly referred to as the "placentals", since all therians possess placentae) (Figure 1). The monotremes, represented by the platypus and two echidna species, are thought to have diverged from the Therian line approximately 170 million years ago (MYA) (Air et al., 1971, Pask and Graves, 1999) and display a mixture of ancestral reptilian traits and derived mammalian characteristics.

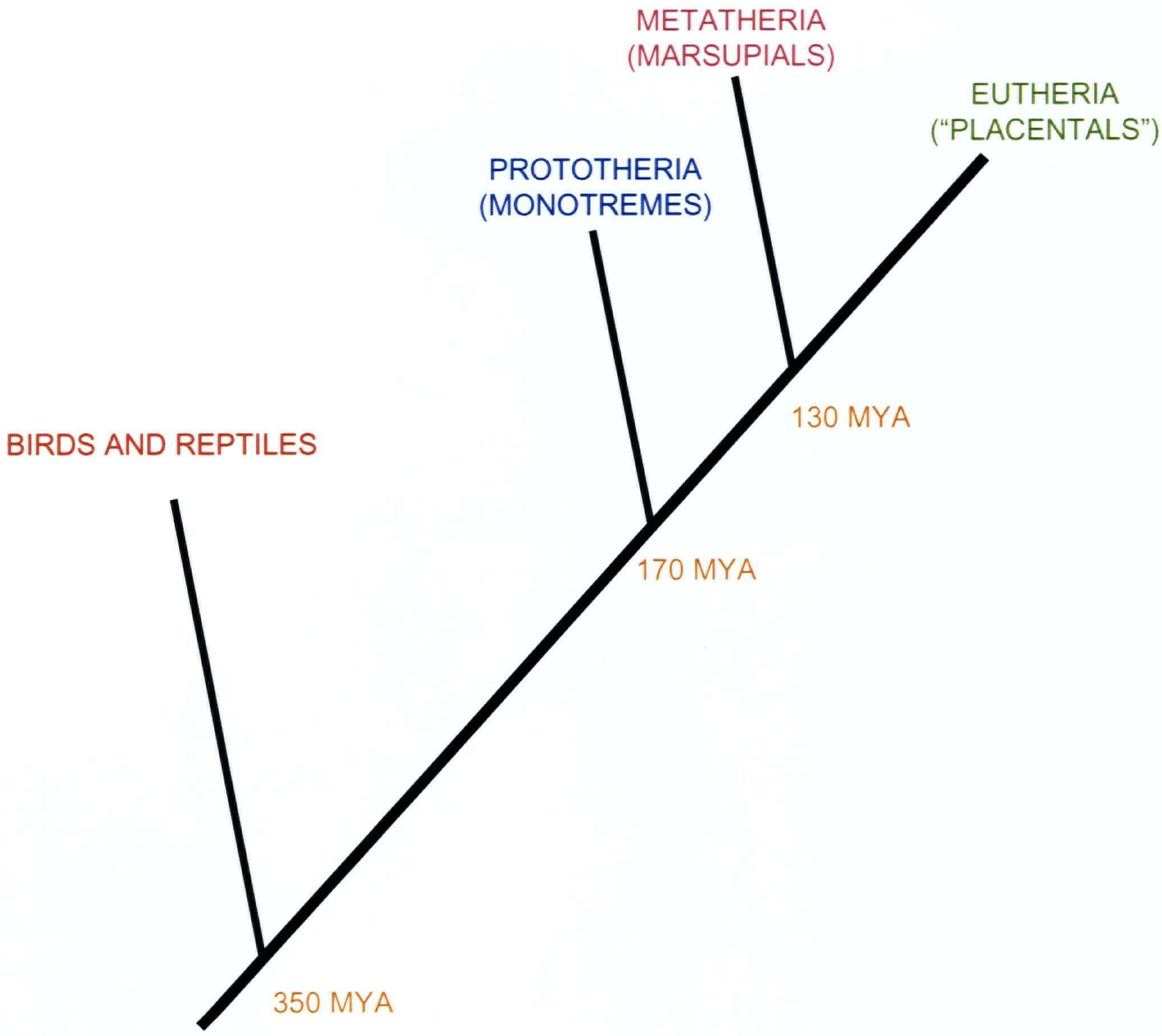


Figure 1: Phylogeny of the Class Mammalia (adapted from Pask and Graves, 1999).

They are ovoparous yet possess hair and milk glands. In contrast to all other mammals, which maintain a constant body temperature of 30-39° C, the internal temperature of a platypus is only about 30° C (Campbell, 1993). All the Therian mammals, which split into the metatherian and eutherian line 130 MYA (Pask and Graves, 1999), are viviparous, but differ in the placental type and the maturity of young at birth. The tiny marsupial embryo exits the uterus only about a month after fertilization and continues its development in the marsupium or pouch. The eutherians, the most numerous and familiar group of mammals, give birth to relatively well-developed live young and possess an allantoic placenta.

While the chromosomal sex determination system, in which XX defines a female and XY a male, is conserved within mammals there are differences in the way primary and secondary sex characteristics are controlled among the Prototheria, Metatheria and Eutheria. Additionally, there are slight differences within the eutherians themselves such as the loss of the Y in mole-voles and multiple gonosome systems in lemmings and Akodon rodents (Solari, 1993), but these organisms seem to represent exceptions to the established eutherian system.

In the vast majority of eutherians the Y chromosome carries a dominant "testis-determining factor" and defines the male phenotype; individuals with an XO genotype are female, while a Y with any number of X chromosomes (XXY, XXXY) will be phenotypically male. In marsupials the Y also controls testis determination, however it seems that the presence of the male gonad is not sufficient for the development of all the secondary male characteristics, as is the case in eutherian mammals. An XXY animal has no scrotum and possesses a pouch with mammary glands, while an XO individual lacks a pouch and mammaries, but has a scrotum (Pask and Graves 1999). The situation in monotremes is not known because the animals do not breed in captivity (Watson et al., 1991).

The eutherian X chromosome is large (165 Mb (megabases) in humans), representing 5% of the haploid genome, and may carry as many as 3000 genes (Strachan and Read, 1999). In contrast, the largely heterochromatic Y chromosome is estimated at a mere 60 Mb in length (in humans) and contains approximately 30 functional genes, some of which have homologues on the X chromosome, which are generally not inactivated, and others which are Y-specific and exclusively expressed in the testis (Strachan and Read, 1999). Both the X and Y can be divided into differential regions and pseudoautosomal regions - areas that are shared between the sex chromosomes and, as a result, do not appear sex-linked. The human gonosome pair possesses two pseuodoautosomal regions (PARs), PAR1 at the tips of the short arms and PAR2 at the tips of the long arms, at which recombination still takes place (Graves et al., 1998b). The gene content and size of the eutherian X chromosome are remarkably conserved, constituting 5% of the haploid genome regardless of the size of the autosomes (Ohno, 1967). Ohno (1967) postulated that this was due to mammalian X inactivation and that any rearrangements of this chromosome would be selected against since they would disrupt dosage compensation.

Unlike the differential region of the X, the size, content and number of the pseudoautosomal regions varies widely between species. Comparisons with metatherian and monotreme chromosomes have shown that the genes in the eutherian PARs are autosomal in these diverged mammals, indicating that there have been several independent additions of autosomal material to both the X and Y in various eutherians (Solari, 1993, Graves et al., 1998b). The PAR genes are not subject to inactivation and their dosage is equal in the two sexes (Graves et al., 1998b). While pairing and recombination at the large PAR in both mice and humans is essential for spermatogenesis and fertility, the difference in gene content between these two species indicates that this function is not sequence-dependent (Graves et al., 1998b). Additionally, in some

rodents and all marsupials the gonosomes do not synapse at all, challenging the notion that XY pairing is essential for meiosis and male gametogenesis (Solari, 1993, Graves et al., 1998b).

Monotreme sex chromosomes are relatively large and the X and Y pair at meiosis over the entire short arm of the X and the long arm of the Y (Pask and Graves, 1999). The marsupial X is smaller than its eutherian counterpart and does not pair with the extremely small Y chromosome (12 Mb) at meiosis (Pask and Graves, 1999). Despite the lack of a pairing PAR, the marsupial gonosomes form an unusual XY body ("balloon" or "dense plate structure") during male meiosis (Solari, 1993). Comparative mapping studies have shown that the region corresponding to the current marsupial X chromosome, which maps to the monotreme X and includes only the genes of the long arm and pericentric region of the human X, may represent the ancestral X chromosome content (Figure 2) (Pask and Graves, 1999). The remainder of the genes on the short arm of the eutherian X, including the large PAR, lie on the autosomes in monotremes and marsupials, indicating that they were a later addition to the eutherian gonosomes. Additionally, Pask and Graves (1999) believe that the large region of homology between the monotreme X and Y is another, independent addition of autosomal material to the sex chromosomes in the Prototheria.

Although the X and Y chromosome pair of extant mammals may differ drastically in their gene content and size, there is ample evidence that the sex chromosomes of modern mammals evolved from an ancient autosomal pair. The strongest evidence comes in the form of homologous X-Y genes present on the differential regions of these chromosomes. The *UBE1X/Y* (ubiquitin-activating enzyme E1) and *SMCX/Y* (selected mouse cDNA on the X) genes, as well as *SRY* and its ancestor *SOX3* (SRY-box containing gene 3), map to the X and Y in eutherians and marsupials (Graves et al., 1998a). Additionally, there are several marsupial X-

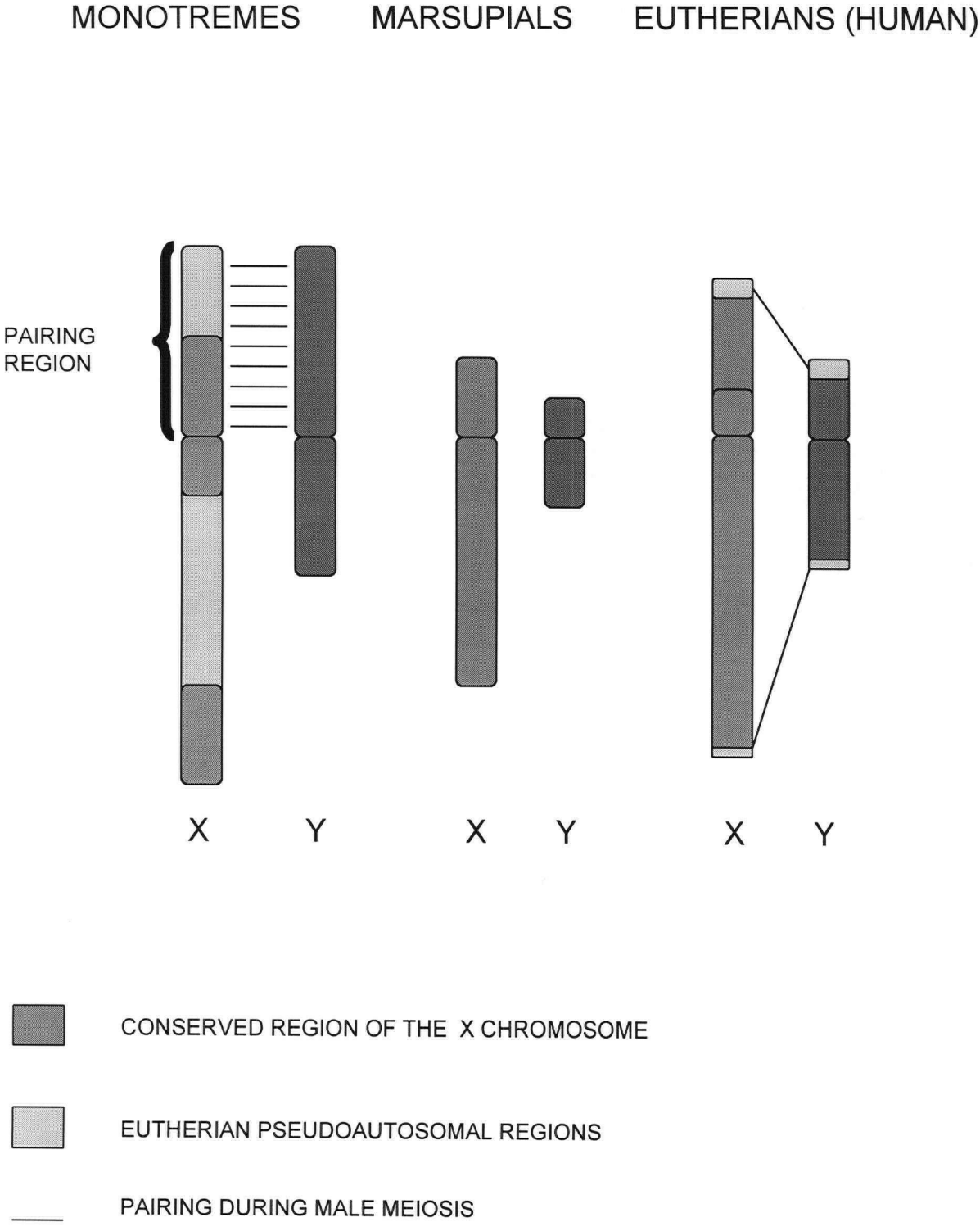


Figure 2. Sex chromosomes in the three mammalian subclasses.

specific genes that map to the monotreme XY pairing region (Watson et al., 1991), indicating an ancient homology.

While the exact sequence of events in the evolution of the sex chromosomes is unknown, it must have involved the acquisition of a dominant male-determining gene on the Y, probably *SRY*, suppression of recombination between the proto-X and proto-Y chromosomes followed by decay of most Y-linked genes and the inactivation of most X-linked genes. An interesting exception is the case of the *SYBL1* (synaptobrevin-like 1) and *HSPRY3* (sprouty (*Drosophila*) homologue 3), two genes located on the smaller PAR of the human X and Y (Lahn et al., 2001). These genes appear to undergo inactivation on both the X and Y chromosomes, suggesting that there may have been a recent X to Y translocation in the human lineage, since they are not present on the Y in other species (Lahn et al., 2001). The differentiation of the sex chromosomes was driven by the lack of recombination of the Y, which led to an accumulation of deleterious mutations as a result of genetic drift (Muller's ratchet), by hitchhiking (being carried along with a beneficial mutation) or the effects of sexually antagonistic genes (Graves et al., 1998a, Strachan and Read, 1999). A study by Lahn and Page (1999) based on the rate of synonymous nucleotide divergence estimated the time of X-Y divergence to be 240-320 million years ago, shortly after the mammalian and avian lineages spilt. The authors were able to divide the human X chromosome into four strata that range from the oldest proto-X material to subsequent additions of autosomal material. According to their research, the oldest gene pair on the human X and Y is the *SRY/SOX3* pair, which is in accordance with the proposed model of sex chromosome evolution. The *XIST* gene, which has a crucial role in X-chromosome inactivation, was found to be slightly younger than the oldest gene pairs, which argues that

dosage compensation evolved after sex chromosome differentiation, although the possibility remains that it was X-inactivation that lead to Y attrition (Graves et al., 1998a).

The end point of Y degeneration may be a complete loss of the chromosome, a process that seems to have already taken place in two species of voles, *Ellobius tancrei* and *Ellobius lutescens*, which lack a Y chromosome and the *Sry* gene (Solari, 1993, Just et al., 1995, Vogel et al., 1998). Both male and female *E. lutescens* carry only one X chromosome, while in *E. tancrei* all individuals possess an XX pair (reviewed in Solari, 1993). A recent study has shown that *SOX9*, an autosomally located gene involved in testis differentiation, is not the testis determining factor in *E. lutescens* (Baumstark et al., 2000), leaving sex determination in these rodents a complete mystery.

A consequence of the mammalian system of dosage compensation is the reduction of X-linked gene products by half in both males and females – a condition that would probably be very deleterious to both sexes. Therefore, it is expected that an increase in the rate of transcription of genes on the X chromosome would accompany the loss of the X in the male and the functional hemizygosity of the female (Ohno, 1967). Adler and colleagues (1997) compared the transcription rate of the *Clc4* (chloride channel 4) gene, which is autosomal in *Mus spretus* and X-linked in *M. musculus*. The activity of the autosomal gene was half that of the X chromosome copy, supporting Ohnos's prediction.

The entire eutherian X, excluding the pseudoautosomal regions, is remarkably conserved within the subclass, but not within the Class Mammalia. At the class level, Ohno's law only applies to the original, conserved proto-X shared by all three mammalian subclasses. Ohno hypothesized that this conservation is due to the phenomenon of X-inactivation, which occurs in both eutherians and metatherians, but it is still unclear why the monotreme X-chromosome

would share this conserved region since these animals may not undergo X-inactivation (Cooper et al., 1993, Graves et al., 1998a, Watson et al., 1990). Graves and colleagues have proposed that the monotreme and marsupial genomes may be more stable than that of eutherians, since the three extant monotreme species, although they have been separated for 50 million years, retain a virtually unchanged karyotype (Solari, 1993). Similarly, the ancestral marsupial karyotype or some simple derivative is shared by virtually all the species (Solari 1993, Graves et al., 1998b). Additionally, recent studies of non-mammalian vertebrate genomes have shown a conserved vertebrate genome even in fish, indicating that the large variations in eutherian karyotypes may be the exception rather then the rule (Graves et al., 1998a).

X-chromosome inactivation in the Eutheria

Although there are many unknowns regarding the mechanism involved in the transcriptional silencing of the X, even among the extensively studied eutherian models, the characteristics of the inactive chromosome are well established. The inactive X chromosome in eutherians is late replicating, hypermethylated at the promoter and CpG islands at the 5' ends of genes, hypoacetylated at the lysine residues of core histones, accumulates a novel histone variant macroH2A (Avner and Heard, 2000) and expresses the *XIST* gene (X Inactive Specific Transcript, *Xist* in mouse). The process of X-chromosome inactivation is thought to consist of three distinct phases: initiation, spread and maintenance (Heard et al., 1997) and is under the control of the X-chromosome inactivation centre (*XIC*). Initiation involves the counting of X chromosomes in the cell and choice of which one to inactivate, or more probably, the choice of

which one will remain active. The inactivation signal then spreads *in cis* from the *XIC*, silencing most of the genes on the chosen X chromosome. Maintenance mechanisms, such as DNA methylation, ensure the inactivation of the chromosome is clonally transmitted to the future cell generations. Based on observations in murine female embryonic stem (ES) cells, *Xist* expression, late replication and gene silencing are the earliest signs of inactivation, occurring at 1.5-2 days of differentiation (Keohane et al., 1998). These changes are followed by histone deacetylation at four days and, finally, DNA methylation at 21 days after induction of differentiation (Keohane et al., 1998).

X-inactivation takes place very early during embryogenesis, approximately at the late blastocyst stage, and is stably maintained within the somatic cells of eutherians (Heard et al., 1997). During oogenesis, however, the silenced chromosome will be reactivated. In the mouse the extraembryonic trophoectoderm initiates inactivation at 3.5-4.5 days post coitum (dpc) and the primitive endoderm (yolk sac) from 4.5 to 5.5 dpc. The embryonic (epiblast) cells do not inactivate until after implantation at day 6.5, at which point the parental imprint has been erased and inactivation is random (reviewed in Lee, 2000). In the somatic tissues of eutherians both the paternal or maternal X can be inactivated. However, in the extraembryonic tissues of some eutherian mammals (mouse, rat) and in the somatic tissues of marsupials it is always the paternal X that is silenced (Avner and Heard, 2000).

Since X-chromosome inactivation is thought to be a means of dosage compensation it is not expected that silencing would take place in the male. Interestingly, transcriptional silencing of the X and *Xist* expression does take place in the male germline during spermatogenesis, albeit at much lower levels than in female somatic tissues (Kay et al., 1993, Ayoub et al., 1997). Both the X chromosome and its pairing partner, the Y chromosome, become condensed and late

replicating at the onset of meiosis and combine to form a sex vesicle or XY body, which has been reported in most eutherian and marsupial species (Solari, 1993). This process is transient and restricted to the period of meiosis, as a number of X-linked genes show post-meioitic transcription in spermatids (Hendriksen et al, 1995). The silencing of the X chromosome during male gametogenesis may be a mechanism to avoid possible damage from illegitimate recombination between the two sex chromosomes or the formation of unpaired recombination sites on the single X chromosome (Heard et al., 1997) and is probably distinct from X-inactivation in the female. Although the XY body accumulates the novel histone macroH2A (Hoyer-Fender et al., 2000) like the inactive X chromosome, the male X lacks the histone hypoacetylation seen in female somatic cells (Armstrong et al., 1997). The most significant difference was detected by Marahrens et al. (1997) who found that male *Xist* null mice were healthy and fertile, indicating that the silencing of the X in the male germline may not be crucial or that it may proceed through an *Xist*-independent mechanism. Another intriguing observation, obtained using an X-linked *lacZ* transgene, is the transient silencing of the X chromosome in the somatic cells of the urogenital ridge in early male mouse embryos (10.5-11.5 dpc) (Jamieson et al., 1997). The authors have suggested that this may inactivate dosage-sensitive sex reversal loci on the X chromosome and may allow *Sry* to more efficiently guide testis differentiation.

The elements of X-chromosome inactivation

Late replication timing

One of the hallmarks of the inactive X, and one of the earliest signs of inactivation, is the appearance of a late-replicating chromosome. Replication late in the S phase is frequently used as a criterion to determine whether X:autosome translocations have undergone silencing. Although the chromosome as a whole appears late-replicating, studies at the single gene level have shown that replication timing corresponds to the transcriptional activity of genes or chromosomal domains (Heard et al., 1997). Genes subject to inactivation will replicate later than their counterparts on the active X chromosome, however, genes that escape inactivation replicate synchronously with the rest of the genome. Demethylating treatments with 5-azacytidine, which lead to chromatin decondensation and earlier replication in gerbil fibroblasts, have implicated DNA methylation in the control of replication timing (Heard et al.,1997).

Histone modifications

Histone modification plays an essential part in mammalian dosage compensation, as it does in *Drosophila*, and is probably an ancient and conserved feature of inert chromatin. Both the constitutive heterochromatin and inactive X chromatin of mammals are enriched for the hypoacetylated forms of histones H2A, H3 and H4 (Heard et al., 1997). The acetylation of histones at specific lysine residues is thought to reduce their affinity for DNA and each other,

leading to a open chromatin structure that is more conducive to transcription (Strachan and Read, 1999). According to this model, hypoacetylation of histones results in a closed chromatin conformation. Recent studies have linked the silencing ability of the methyl-binding protein MeCP2 to the recruitment of histone deacetylases (Ng and Bird, 1999). Additionally, the maintenance methyltransferase, DNMT1, can interact with histone deacetylases and repress transcription (Robertson and Wolffe, 2000). These interactions indicate that histone acetylation patterns might target methylation and that, conversely, DNA methylation can attract histone deacetylases through MeCP2.

Until recently it was not possible to examine the histone acetylation status of individual genes and the globally deacetylated status of the X chromosome was inferred from immunofluorescence studies (Heard et al., 1997). This large scale deacetylation happens after four days of differentiation in female ES cells and does not represent an early event in the X-inactivation process (Keohane et al., 1998). However, recent work using a chromatin immunoprecipitation technique has shown that, contrary to the established view, the levels of histone acetylation were comparable on the active and inactive X chromosome. A significant difference in acetylation status was only seen in the promoters of inactivated genes, which showed a dramatic increase in hypoacetylated histones and correlated with the methylated status of the adjacent CpG islands (Gilbert and Sharp, 1999). Since deacetylation precedes methylation, at least in female ES cells (Keohane et al., 1998), this promoter-specific deacetylation may not be due to the deacetylase-recruiting capabilities of MeCP2, but may precede it (Gilbert and Sharp, 1999, Ng and Bird, 1999).

More support for a crucial role for histone deacetylation in the process of X-inactivation comes from experiments in which cells were treated with the deacetylase inhibitor trichostatin A

(TSA). The treated cells not only had increased levels of acetylated histones, but may also be defective in *Xist* localization and gene silencing (O'Neill et al., 1999). O'Neill and colleagues also reported the existence of a differentially hyperacetylated region upstream of the main *Xist* somatic promoter (1999). The hyperacetylated domain is only present in undifferentiated female ES cells and disappears upon differentiation, indicating a role in the choice of chromosome to be inactivated, possibly facilitating the upregulation of *Xist* expression through the maintenance of an open chromatin conformation (O'Neill et al., 1999). Since male ES cells and female cells with only one functional *Xist* allele do not show this hyperacetylation mark, the authors suggest it may also have a role in chromosome counting. The combined results of these recent experiments (Gilbert and Sharp, 1999, O'Neill et al., 1999) imply that histone modifications may play a crucial role not only in the maintenance of the silenced state, but also in the initiation of the X-chromosome inactivation.

Another role for histones in X-chromosome inactivation was revealed by the discovery of a novel histone variant, macroH2A, which associates with the inactive X chromosome to form a macrochromatin body (MCB) in humans, mice and dogs (Costanzi and Pehrson, 1998). Conditional deletions of the *Xist* gene in mouse embryonic fibroblasts result in the disruption of the MCB, indicating that the *Xist* RNA is necessary for histone macroH2A deposition on the inactive X chromosome. However, the lack of both the *Xist* RNA and the MCB do not affect X silencing once it has been established and will not lead to the reactivation of the chromosome (Csankovski et al., 1999). Previous results had indicated that *XIST* is not associated with the nucleosome, as it remained attached to the nuclear matrix when the chromosomal DNA and histones were removed (Clemson et al., 1996). However, using a more sensitive method Gilbert et al. (2000) showed that *XIST* RNA can be co-precipitated by antiserum for macroH2A

and by antisera that recognize hypoacetylated histones H3 and H4. This discrepancy can be explained by the fact that the older study used relatively harsh conditions when treating the nuclei, which may have lead to the disruption of the putative *XIST*-histone complexes (Gilbert et al., 2000). These results, along with an indication that the novel domain of macroH2A may have an RNA-binding function (Pehrson and Fuji, 1998), support a role for *XIST/Xist* in chromatin remodeling, possibly by nucleating ribonucleoproetin chromatin modifying complexes as is the case in *Drosophila* (Gilbert et al., 2000, Kelley and Kuroda, 2000a, b).

Hypermethylation

Although the global methylation levels on the inactive and active X chromosomes seem to be comparable, methylation of CpG islands at the 5' end of genes on the inactive X chromosome is thought to be a crucial factor in the stable maintenance of inactivation (Heard et al., 1997). Methylation patterns on the inactive X chromosome clearly differ between individual genes and correlate with expression – genes that are transcriptionally repressed are methylated at the CpG islands, while those that escape X-inactivation are not (Heard et al., 1997, Hansen et al., 1998). Lack of such methylation on the marsupial X chromosome (Kaslow and Migeon, 1987) and hypomethylation in the extraembryonic tissues is thought to be the reason for the relative instability of silencing in these two systems (Heard et al., 1997). Despite its late appearance in the silencing process, methylation has been the prime candidate for a epigenetic mark that would be involved in counting, chromosome choice or imprinted X-inactivation in the extraembryonic tissues, based on its crucial role in the regulation of

autosomal imprinted genes (Heard et al., 1997, Sado et al., 2000, Lee, 2000). However, as of yet, no differential methylation has been identified that fulfills the criteria of an epigenetic mark involved in the initiation of the inactivation process. The majority of the evidence indicates that methylation may not be an initial event of X-inactivation, but a later stabilizing mechanism (Heard et al., 1997).

The X Inactivation Centre (XIC)

The *XIC/Xic* (human/mouse), defined as an 80 kilobase (kb) fragment in the mouse and a one megabase region in humans, is located on the X chromosome and contains several elements involved in the inactivation process (reviewed in Heard et al., 1997, Avner and Heard, 2001). It is essential for silencing to occur *in cis*, and an X chromosome lacking the *XIC* cannot be inactivated. This region is also thought to be necessary for the counting of the ratio of X chromosomes to autosomes. The putative counting mechanism is unknown, although the most favoured hypothesis is that of an autosomal blocking factor or nuclear attachment site that would be present in limited quantities and bind to only one X chromosome per autosomal complement. While this mechanism remains elusive, it is clear that it exists, as cells that are aneuploid for the X (47, XXX and 48, XXXX) inactivate all but one chromosome (Brown and Willard, 1993). Tetraploid cells contain two inactive and two active X chromosomes, while the situation in triploids is less clear and 69, XXX and 69, XXY triploids have been found to have one or two active X chromosomes (Brown and Willard, 1993).

Although X-chromosome inactivation is random in eutherian somatic tissues, in the absence of selection against a particular X at the cellular level, a locus named the X controlling element

(*Xce*) was found to influence the probability that a particular chromosome would be silenced (Avner and Heard, 2000). A chromosome with a "strong" *Xce* allele, contained within the *Xic* in mouse, but distinct from *Xist*, would remain active more often than one with a weaker allele (approximately 70%:30% ratio).

The XIST/Xist gene

In recent years significant advances have been made in the understanding of X-chromosome inactivation, however, the function of the most significant element in the silencing process, the *XIST/Xist* gene, is still a mystery. The *XIST/Xist*[1] gene is contained within the *XIC/Xic* and is the only gene expressed exclusively from the inactive X chromosome, resulting in a large (17 to 19 kb in human, 15 to 18 kb in mouse) spliced, polyadenylated transcript (Brown et al., 1991, Brockdorff et al., 1992, Brown et al., 1992, Hong et al., 2000). The *XIST/Xist* RNA is not translated and is retained in the nucleus where it associates, through a yet unknown mechanism, with the chromosome from which it originated (Clemson et al., 1996). *XIST/Xist* expression is thought to be an early, possibly the primary, step in the inactivation process and chromosomes lacking the gene cannot inactivate (Penny et al., 1996, Avner and Heard 2001). The *XIST/Xist* RNA coating the inactive X dissociates from the chromosome as the cells enter mitosis and is re-transcribed in the daughter cells (Lee and Jaenisch, 1997).

Based on an experiment in which a transgene carrying *Xist* and 15 kb of surrounding sequence was sufficient for inactivation, Herzing et al. (1997) have even suggested that *Xist*

[1]By convention, the human X Inactive Specific Transcript gene is capitalized (*XIST*), while homologous genes in other species are not (*Xist*).

may be the *XIC* and is the only sequence necessary for silencing. However, various transgenes, females with deletions downstream of *XIST* and mice with induced deletions 3' of *Xist*, suggest that the counting and choice function of the *XIC/Xic* lies outside of *XIST/Xist* (reviewed in Avner and Heard, 2001, Wutz and Jaenisch, 2001).

Penny et al. (1996) demonstrated, through the use of a deleted *Xist* allele in which the minimal promoter was left intact, that the presence of the *Xist* transcript, rather than just the act of transcription in the region of the gene, is essential for silencing to occur. An upregulation of the *Xist* gene and increased transcript stability seems to be the trigger for inactivation (Avner and Heard, 2001, Wutz and Jaenisch, 2001). While some authors (Johnston et al., 1998) have suggested that this increased half-life is due a promoter switch, those results have not been substantiated (Warshawsky et al., 1999). An alternative theory holds that the increased RNA stability depends on the accumulation of a threshold amount of the transcript and the instability of *Xist* in normal ES cells may be a result of poor expression (Wutz and Jaenisch, 2001). It is still unclear what leads to the increase in transcription, but a differentially acetylated region upstream of the promoter is a candidate (O'Neill et al., 1999). The histone H4 hyperacetylation, which is seen only in undifferentiated female cells and disappears upon differentiation, may lead to transcriptional upregulation through the maintenance of an open chromatin conformation. Additionally, the *Tsix* antisense transcript seems to be involved in *Xist* expression in both random and imprinted inactivation (see below).

When the *XIST/Xist* promoter is methylated, as it is on the active X chromosome, the gene is not expressed (Brown, 1992, Norris et al., 1994, Beard et al., 1995). In mouse embryos deficient for the DNA methyltransferase gene, *Xist* is hypomethylated and expressed in both males and females (Heard et al., 1997). This expression may simply be due to the global

hypomethylation of the genome in these cells, rather than a result of the specific loss of *Xist* methylation. Initial results indicating that parental differences in the methylation of the promoter could be responsible for the imprinted inactivation seen in extraembryonic tissues (Ariel et al., 1995, Zucotti and Monk, 1995) have been contradicted by more recent experiments that have shown that the parental alleles do not exhibit differential methylation in the region (McDonald et al., 1998). Thus, the methylation pattern seen in the promoter and first exon of *XIST/Xist* seems to be a result of its transcriptional status.

Experiments involving transgenes containing the *XIST/Xist* genes and surrounding sequence have shown successful transcriptional silencing of autosomes into which they have integrated, even in male cells (Migeon et al., 1999), and confirm the observations from X:autosome translocations, which show that some autosomes are more amenable to the spread of inactivation than others (Lyon, 1998). Interestingly, once the inactivated state is established, as it is somatic cells, deletion of *XIST* or the *XIC* does not seem to lead to reactivation (Brown and Willard, 1994). Likewise, reactivation of *XIST* through demethylation in rodent-human somatic cell hybrids, while it leads to *XIST* expression, does not lead to the correct *XIST* localization nor X-inactivation (Hansen et al., 1998, Tinker and Brown, 1998). Subsequent experiments have indicated that even with *Xist* expression and correct localization, X-chromosome inactivation cannot be induced in adult somatic cells, indicating that the developmental timing of *XIST* expression is crucial (Clemson et al., 1998).

Experiments by Wutz and Jaenisch (2001) involving an inducible *Xist* cDNA transgene in male ES cells support these observations. The authors found that they could induce reversible *Xist*-dependent repression only in undifferentiated ES cells. Once the cells had begun to differentiate, however, the inactivation became irreversible and independent of ongoing *Xist*

expression. The transgenic *Xist*, although it lacked the counting function, led to late replication and histone hypoacetylation of the inactive chromosome once the cells had differentiated. An intriguing observation, which may shed light on the mechanism of *Xist* function, was the *Xist*-induced silencing in undifferentiated ES cells that proceeded without any detectable change in replication timing or histone hypoacetylation.

There is no doubt that the *XIST/Xist* RNA is an essential part of eutherian X inactivation, yet it does not exhibit a high degree of sequence conservation. The only two genes to have been sequenced in their entirety, the mouse and human homologues, exhibit only 70% sequence identity (Brockdorff et al., 1992, Brown et al., 1992). A study by Hendrich and colleagues (1993) examined the evolutionary conservation of the gene in a number of mammals (several primates, fox, cat, rabbit, cow, hamster and horse). By performing Southern blots with probes from various regions of the human *XIST*, the authors were able to identify a region of increased sequence homology in the 5' region of the gene. Sequencing of a subset of the animals revealed a series of eight to nine repeats, encompassing approximately one kilobase, with a highly conserved core. However, alignments of the surrounding sequenced regions indicated a modest homology. The level of conservation was less than what is seen in protein-coding exons, yet significantly higher than found in introns, suggesting a selective pressure. Since *XIST/Xist* probably functions at the RNA level, appropriate secondary structure, rather than strict sequence conservation, would be the expected target of natural selection. Additionally, methylation analysis in the putative promoter region of the gene showed a similar methylation pattern in human, murine and bovine samples. While the authors note that the 5' conserved repeats represent the strongest candidate for a functional domain, this hypothesis has not been confirmed. Despite the low conservation, crucial regions seem to be conserved between the

mouse and human homologues since a human *XIST* YAC has been shown to induce inactivation in mouse autosomes, although without the accompanying late replication and histone hypoacetylation (Heard et al., 1999, Migeon et al., 1999). The absence of histone modifications and late replication in this system, along with the observation that in hamster-human somatic cell hybrids *XIST* is not properly localized and cannot induce repression, point to the existence of species-specific autosomal factors that contribute to the inactivation process (Hansen et al., 1998).

Although evidence is currently lacking, most authors believe that *XIST/Xist* acts as part of a ribonucleoprotein complex that is capable of chromatin remodeling and gene silencing (Mlynarczyk and Panning, 2000) in a manner analogous to the *roX* RNAs in *Drosophila* (reviewed in Kelley and Kuroda, 2000a,b). This theory is supported by the co-immunoprecipitation of *Xist* and hypoacetylated histones H3 and H4, and the requirement of *Xist* for the deposition of macroH2A, containing a putative RNA-binding domain, on the inactive X chromosome (Pehrson and Fuji, 1998, Gilbert et al., 2000). Both *XIST/Xist* and the *roX 1* and *roX 2* RNA are expressed from the X chromosome and spread *in cis* from their site of synthesis, although the extent of *Xist* spread is much greater, and both involve a change in histone acetylation. All three RNAs are spliced and seem to coat the X chromosome in a banded pattern (Duthie et al., 1999, Kelley and Kuroda, 2000a,b). While both systems involve a non-coding RNA expressed from the X chromosome and are the only RNAs to spread from their point of origin, no indications of a common evolutionary origin have been reported (Kelley and Kuroda, 2000b). The *roX* RNAs associate with a six-protein complex that mediates male-specific hypertranscription of the X-linked genes. Mutations in five of the genes forming this group, known as the male-specific-lethal (MSL) complex, have been isolated and all cause male

death. Five of the protein components have been characterized so far and fall into two categories. The MLE component is an RNA helicase, the MOF is a histone H4 acetyltransferase, while JIL-1 can phosphorylate histone H3 in vitro. The remaining two proteins seem to bind RNA through their chromodomains. The MSL group seems to recruit nascent *roX* RNA at two chromatin entry sites to form a mature complex. Unlike *XIST/Xist*, which only coats its chromosome of origin, the *roX* RNAs seem to be able to locate the X *in trans* and appear to diffuse freely between the 35 chromatin entry sites. Not much is known about the sex-specific regulation of *roX* expression, although there are no indications that an antisense RNA, analogous to *Tsix,* is involved (see below).

Tsix

Another crucial element of the *Xic*, at least in the mouse, seems to be the *Tsix* antisense RNA, which is transcribed in undifferentiated ES cells and early embryos and spans the entire *Xist* gene (Lee et al., 1999). Very early in embryogenesis low levels of the *Xist* and *Tsix* RNA are expressed from every *Xic* present. The *Xist* RNA is unstable during this pre-inactivation stage and does not advance from its site of transcription (Mlynarczyk and Panning, 2000). Once inactivation is initiated, through an unknown mechanism, *Xist* RNA is seen to spread over the chromosome that will be silenced, while *Tsix* expression stops. On the remaining X chromosome, low levels of both *Xist* and *Tsix* transcripts persist for some time and then cease. The 5' end and promoter of *Tsix* lie within the *DXPas34* locus - a short CpG rich region that is hypermethylated on the active X chromosome. The degree of methylation at this locus seems to

correlate with the tendency of a chromosome to avoid inactivation and is reminiscent of the *Xce*, although mapping studies show the two loci to be separate (Avner and Heard, 2000). Deletions of the 5' end of *Tsix* and *DXPas34* lead to the exclusive inactivation of the deleted chromosome in ES cells, indicating that these elements influence X-chromosome choice and imprinting, but not counting.

Recent research has shown that *Tsix* is not only involved in random silencing, but is vital for imprinted inactivation in mouse extraembryonic tissues (Lee, 2000). In normal mouse embryos the maternal chromosome is thought to be resistant to the inactivation signal or shows a higher affinity for an activating factor, and it is the paternal chromosome that is silenced. Sometime after the differentiation of the extraembryonic lineages, the parental mark is lost and inactivation is random in the embryo proper. Mice inheriting a *Tsix* gene deletion at the 5' CpG island from their fathers were born at the expected frequency, while there was a marked shortage (18% survival rate) of pups carrying the deletion on the maternal X chromosome, indicating a severe parent-of-origin effect. In the mutant male embryos the deleted maternal X was inactivated, while both the paternal and maternal X chromosomes were silenced in mutant female embryos. This embryonal lethality could be due to the inability to reactivate the deleted X chromosome during oogenesis or to the failure of imprinted inactivation in the extraembryonic tissues. Lee (2000) found that the cause of death was due to the incorrect inactivation of the maternal X chromosome in the extraembryonic tissues, rather than a failure in oogenesis. These results are consistent with previous work, which showed that the maternal X chromosome in rodent extraembryonic tissues is normally not silenced (Marahrens et al., 1997, Goto and Takagi, 1999). Lee proposes that the probable imprint differentiating between the maternal and paternal chromosome, with respect to *Tsix* expression, is the differentially methylated CpG region of

DXPas34. Studies on other imprinted genes have shown that the putative imprinting centres contain differentially methylated CpG islands and are located within or near genes coding for antisense and noncoding RNAs (reviewed in Lee, 2000). Abolition of methylation through targeted knockouts of the *Dnmt 1* maintenance methyltransferase results in the loss of imprinted gene expression (reviewed in Robertson and Wolffe, 2001). The identification of differential methylation in the 4kb CpG-rich putative promoter and transcriptional start site domain of *Tsix*, also known as the *DXPas34* region, strengthens this hypothesis (Courtier et al., 1995).

However, more recent experiments by Prissette and colleagues (2001) have not revealed a differential methylation pattern in extra-embryonic tissues by bisulfite sequencing. Additional evidence that methylation may not be the imprint is given by the experiments of Sado and colleagues (2000) in which X-inactivation was examined in mice lacking the maintenance methyltransferase *Dnmt1*. Previous work had shown that targeted disruption of *Dnmt1* led to the lethal ectopic *Xist* expression in male embryos and ES cells, and expression from both X chromosomes in female ES cells. This expression was presumably due to the loss of methylation at the 5'region of *Xist* and it was thought that it may lead to inappropriate X-inactivation *in vivo* (Beard et al., 1995, Panning and Jaenisch, 1996). However, the results of Sado et al (2000) show that the lack of *Dnmt1* does not seem to disrupt imprinted X-inactivation in the extraembryonic tissues of the mutant mice embryos indicating that methylation may not be the imprinting mark in this case (2000). However, they report a disruption of inactivation and reactivation of the silenced X chromosome in the embryo proper, supporting a role in the maintenance and stabilization of X-chromosome inactivation. The authors hypothesize that histone deacetylation may have a dominant role in the imprinted silencing seen in extraembryonic tissues and that histone deacetylases, which have a universal role in gene

silencing regardless of whether an organism possesses methylation, may have a more fundamental role than previously thought. Although a differentially hyperacetylated region was identified upstream of the *Xist* promoter in undifferentiated female ES cells, it has not been examined in extraembryonic tissues (O'Neill et al., 1999). Further work needs to be done to determine whether a sex-specific mark, possibly due to differential hypoacetylation, is responsible for the inactivation of the paternal X chromosome in rodent extraembryonic tissues.

In conclusion, the *Tsix* deletion in ES cells leads to the virtually exclusive inactivation of the mutant chromosome. In extraembryonic tissues the parental imprint overrides the counting mechanism and leads to paternal inactivation in the vast majority of cells, despite the fact that the maternal chromosome is already inactivated due to the *Tsix* deletion. The combined evidence points to a role for *Tsix* in chromosome choice and imprinting, but not in counting. However, neither the inactivation of the paternal X (Xp) in ES cells, nor the lethality of the deleted Xp Xm (maternal) and XmY embryos was absolute. Similarly, XpO mice, although growth-retarded and rare, are viable (Solter and Wei, 1997). These results imply that imprinting must not be complete in extraembryonic tissues, a conclusion supported by previous research estimating that only 85%-90% of rodent extraembryonic cells actually show paternal X-inactivation (Tagaki and Sasaki, 1975, Lee, 2000). Furthermore, since male mice carrying an *Xist* deletion were fertile (Marahrens et al., 1997) and females with a paternal *Tsix* deletion had normal oogenesis (Lee, 2000), it seems that X-inactivation and reactivation in gametogenesis may be proceed through a *Xist-Tsix* independent mechanism or may not be as crucial as in other tissues.

Although expression downstream of human *XIST*, in the putative region of a human *Tsix* equivalent, has been detected, it does not seem to result in the large transcript seen in mouse (J.

Chow, personal communication). Due to the difficulty inherent in studying early human development it remains unclear if a *Tsix* homologue exists, or is necessary, in humans. Unlike mice, humans with supernumerary maternal X chromosomes have relatively normal development (XXX females and XXY Klinefelter males). Most research indicates that humans do not possess a parental imprint on one of the X chromosomes and that either can be inactivated. However, recent results by Looijenga et al. (1999) have shown heterogenous expression of the androgen receptor (ARA) gene in human placenta. Some placentae exhibited preferentially maternal inactivation, some preferentially paternal, while others had biallelic expression. While these observations may be due to some sort of unstable imprinting, they may just be a result of variable inactivation seen for several X-linked genes.

While it seems apparent that *Tsix* is a negative regulator of *Xist*, the mechanism by which this regulation is achieved is unknown. The proposed mechanisms can be divided into two classes: those that require the *Tsix* transcript and those in which the act of transcription is a byproduct of *Tsix* function. In the first group of mechanisms interaction between the *Tsix* and *Xist* transcripts could result in the destabilization of *Xist* secondary structure or might target the putative *Tsix:Xist* RNA duplex for destruction by an RNA interference mechanism (Mlynarczyk and Panning, 2000). In the second class of mechanisms the very act of *Tsix* transcription could regulate *Xist* expression, as is the situation in the imprinted *H19-Igf2* gene pair which are transcribed in a mutually exclusive fashion. The *Tsix* RNA, like the *H19* transcript, may be non-functional (Mlynarczyk and Panning, 2000). Differential methylation of an imprinting centre allows an enhancer element, necessary for both *H19* and *Igf2*, to act only on one of the genes depending on the chromosomal parent-of-origin. In contrast to what was initially thought about *Xist* regulation, the available results indicate that it is the differential epigenetic mark of the

Tsix, rather then the *Xist*, promoter which may be responsible for the imprinted inactivation in the extraembryonic tissues of eutherian mammals.

Spread of inactivation

Although X:autosome translocations are capable of inactivation, it is clear that the spread of inactivation is not as extensive as it would be on the X chromosome (Heard et al., 1997). Additionally, some autosomes and translocations seem more amenable to inactivation than others. Mary Lyon (1998), drawing on the earlier Gartler-Riggs model (1983), proposed that an explanation for this phenomenon could be the presence of certain repeated elements that are ubiquitous in the genome, but are enriched on the X chromosome. These repeats would act as "booster elements" or "way stations" allowing the inactivation signal to spread more efficiently. A good candidate for these repeat elements are the LINE 1 sequences (long interspersed repeat elements), which are particularly concentrated on the X chromosome in both humans and mouse (Lyon, 1998).

LINE elements are retrotransposons that tend to cluster in dark G-bands and are present in all three subclasses of mammals, unlike other repeat candidates, which are specific to certain groups (for example, *Alu* elements are primate-specific) (Lyon, 1998). More recent research by Bailey et al. (2000) using available genomic sequence confirmed that the LINE 1 composition of the human X chromosome is distinct from that of the autosomes and that the highest concentration of LINES is found in Xq13-Xq21, which is the location of the human *XIC*. Furthermore, regions of the X chromosome carrying genes that escape inactivation tended to

harbour less L1 elements than regions that were silenced. The authors also found that the greatest L1 enrichment on the human X was due to a particular subset of eutherian-specific LINEs, which may account for the more stable silencing process in eutherians in comparison with marsupials.

Another tantalizing clue regarding the presence and identity of "booster elements" is the observation that *Xist* RNA coating the inactive X in the mouse, rat and vole appears to have a banded appearance (Duthie et al.,1999), although the pattern seemed to correlate with the LINE-poor G-light bands. This observation may indicate that *Xist* does not directly associate with the repeats, but with the gene rich portions of the chromosome. Lyon (1998) proposes that the LINE elements may aid the spread of inactivation through repeat-induced silencing, a putative cell-defense mechanism present in most eukaryotes. In this model the *XIST/Xist* RNA would act to bind and bring the LINE elements into contact with each other, thus masking intervening non-repetitive sequences and triggering silencing.

Genes escaping inactivation

Despite the various factors involved in inactivation, a number of genes on the X chromosome manage to escape silencing. A recent study has determined that 34 genes out of 224 X-linked genes studied, three of which are pseudoautosomal, escape inactivation in humans (Carrel et al., 1999). The genes that avoid inactivation have acetylated promoter regions (Gilbert and Sharp, 1999), are not methylated at their CpG islands and do not replicate late (Heard et al., 1997). Although genes that escape inactivation are dispersed throughout the human X

chromosome, a significant number are located around the large PAR (Heard et al., 1997). Since X-inactivation is thought to be a means of dosage compensation, it is logical that genes on the pseudoautosomal regions are not inactivated. Likewise, genes with functional Y homologues are not expected to be silent. However, a large proportion of genes adjacent to, but not part of the PAR escape inactivation in humans. This failure to inactivate may be a result of the relatively recent autosomal origin of most of the Xp. Various authors have suggested that a portion of the Xp sequences have not had time to acquire the putative repeat elements necessary for the propagation of the inactivation signal and that they behave more like autosomal material, which is less conducive to silencing (Lyon, 1998, Carrel et al., 1999, Bailey et al., 2000).

An additional evolutionary problem is presented by the several genes with no detectable Y homologues that also escape inactivation, since the presence of Y-linked copy is a strong indicator that a gene will escape inactivation (Jegalian and Page, 1998). It is possible that the recent loss of the Y locus has not allowed sufficient time and selective pressure to include the X-linked homologue into the dosage compensation machinery (Graves et al., 1998b). Carrel et al. (1999) have suggested that the presence of these genes may indicate that strict dosage compensation may not be necessary for all genes on the chromosome. The existence of genes, which are expressed from the inactive X chromosome in only some women, such as the Tissue Inhibitor of Metalloproteinases 1 gene (*TIMP1*) (Anderson and Brown, 1999), along with the observation that some genes in mouse reactivate with age, support this hypothesis (Gartler and Goldman, 1994, Brown et al., 1997). Not surprisingly, the genes that were shown to reactivate in older mice, did not have extensive CpG islands (Gartler and Goldman, 1994).

In contrast to the situation on the human X chromosome, there seem to be remarkably few escapees in the only other well-studied eutherian - the mouse. This difference may be

attributable to limited number of genes that were examined in mouse. The results of Carrel et al. (1999) indicate that the position of a genes (Xp versus Xq) has a statistically significant influence on whether a gene is inactivated or not, and since there are many rearrangements separating the mouse and human X, one would not expect that genes escaping in one species will do so in another. However, the phenotype of an XO (39,X) mouse is much less severe than that in humans with Turner syndrome (45,X) supporting the notion that fewer genes escape inactivation in the mouse. While females with Turner syndrome show a greatly reduced viability *in utero* and a number of phenotypic abnormalities, their murine counterparts only show a slight reduction in developmental viability and fertility (Banzai et al., 1995).

Silencing in the Methatheria and Prototheria and theories about the origins of X-inactivation

While many aspects of X-chromosome inactivation in eutherians remain a mystery, the situation in the marsupial and monotreme subclasses presents an even greater unknown. Late replication and histone hypoacetylation, both hallmarks of inactive chromatin, are conserved in the Metatheria (Cooper et al., 1993, Wakefield et al., 1997). However, marsupial X-inactivation differs from the eutherian model in several crucial ways. In the somatic and extraembryonic tissues of marsupials it is always the paternal X that is silenced, but a Barr body is observed in only a few species (Solari, 1993). Furthermore, this inactivation is often partial and genes may reactivate in particular tissues. Methylation of CpG islands on the inactive X in marsupials has not been detected by restriction enzyme digestion nor bisulfite sequencing (Kaslow and Migeon, 1987, Loebel and Johnstone, 1996). This lack of methylation as a maintenance system has been

hypothesized to be one of the reasons that inactivation in this subclass is unstable. The imprinted nature and instability of silencing in these ancient mammals may represent the ancestral state of X-inactivation.

Perhaps the most significant difference is the apparent lack of an *XIST* homologue in the metatherians, despite intense efforts to locate one (Wakefield et al., 1997, Wakefield, 1998). The investigation of RNA secondary structure identified only two small regions of conservation between the human and mouse genes. In addition to the 5' conserved repeat region identified by Hendrich et al. (1993), Wakefield located another 94 base pair structure in exon four that seemed conserved at the secondary structure level between mouse and humans. The general lack of conservation led the author to speculate that a large portion of *XIST/Xist* may not be crucial for its function, is capable of performing the same function with a very different sequence and structure, or interacts with very different species-specific factors (Wakefield, 1998). Extensive screening of genomic and cDNA libraries from Australian and South American marsupials using a variety of probes from the two conserved regions proved unsuccessful, as were PCR screens involving degenerate oligonucleotide primers designed from the same sequence regions (Wakefield, 1998). Given the crucial importance of *XIST/Xist* in eutherian inactivation, this absence is truly baffling.

X-chromosome inactivation in monotremes is the least studied of all three subclasses due to the fact that these animals do not breed in captivity and their cells are very difficult to culture. The monotreme sex chromosomes, as discussed earlier, share a large region of homology and the short arm of the X is thought to correspond to the marsupial and conserved eutherian X chromosomes. In addition to the incomplete differentiation of the gonosomes, the X and Y chromosomes are involved in a multivalent translocation chain with several small chromosomes

that are unpaired in both sexes (reviewed in Solari, 1993). The only evidence that the Prototheria undergo silencing at all is indirect. Wrigley and Graves demonstrated replication asynchrony of the short, but not the long arm in echidna lymphocytes (1988). This replication timing difference could not be duplicated in fibroblasts, indicating that the inactivation, if it exists at all, is tissue specific. Lack of markers did not allow the authors to determine the parental origin of the late replicating chromosome.

Significant differences exist between the imprinted silencing of the paternal chromosome found in eutherian extra-embryonic lineages and marsupial somatic tissue and random X-inactivation present in eutherian somatic cells. Hypermethylation of CpG islands, an indicator of silencing in eutherians, is not seen in imprinted inactivation and may account for its relative instability (Kaslow and Migeon, 1987). Furthermore, a counting mechanism that determines the sex chromosome:autosome ratio and ensures that only one X per autosome set is active, is only present in random inactivation. In imprinted inactivation, all the paternal and none of the maternal X chromosomes will be inactivated, irregardless of the X:autosome ratio. The mark that protects the murine maternal chromosome from inactivation seems to lie within a CpG rich domain of *Tsix* and was thought to consist of a differential methylation pattern (Lee, 2000), although that is very much in doubt in light of the findings of Prissette et al. (2001). These findings raise interesting questions about the similarities of eutherian and marsupial imprinted inactivation, since marsupials do not seem to methylate X-linked sequences, but do employ histone modification (Kaslow and Migeon, 1987, Wakefield, 1997). However, it is possible that the eutherian parental imprint is in the form of differential acetylation (O'Neill et al., 1999, Sado et al., 2000).

The exclusive occurrence of imprinted X-inactivation in marsupials has lead to the prevailing view that this may represent the ancestral form of mammalian X-chromosome silencing. However, based on new data regarding the origins of autosomal genomic imprinting Ohlsson, Paldi and Graves (2001) investigated the possibility that random inactivation may in fact represent the original situation. Autosomal imprinting, in which a gene's expression is dependent on the parent of origin, shows remarkable similarities to X-chromosome inactivation that indicate a common evolutionary origin. In both processes the silenced allele or chromosome is late replicating, hypermethylated and hypoacetylated. Furthermore, both phenomena result in functional hemizygosity and may involve a differentially methylated inactivation centre that produces an untranslated RNA associated with an antisense transcript. The authors postulate that imprinting arose from random monoallelic expression, which by introducing phenotypic variation may have presented a selective advantage, followed by the skewing of the inactivation process by the appearance of a parental mark. They hypothesize that monoallelic expression began as a stochastic event due to the low transcriptional activity of some genes. In this stochastic inactivation scenario, some cells would have monoallelic expression, others biallelic, while a number would be expected to be functionally null for the gene in question. Eventually, a counting mechanism ensuring proper gene dosage would lead to counted random inactivation. Imprinted inactivation would develop from the random model possibly due to the intrinsic evolutionary competition between the maternal and paternal genomes.

An example of stochastic inactivation is found in the *IL2, IL3, IL4, IL5*, and *IL13* cytokine genes expressed from one or both parental alleles in T-lymphocytes or in the olfactory receptor genes. This stochastic inactivation becomes epigenetically fixed producing a stable monoallelic

expression pattern in the lymphocyte clones. New evidence showing that imprinted *Igf2* gene is transcribed randomly or biallelically in some mouse fetal liver cells, hydatidiform moles and human tumours, indicates that random expression may underlie the normally imprinted status of this gene.

X-chromosome inactivation, however, affects an entire chromosome and almost certainly arose as a means of dosage compensation. The evolutionary progression from stochastic through counted random to imprinted inactivation could not be tolerated since ensuring proper dosage, through counting or imprinting, must have been present from the very beginning. In combination with the complete absence of a random mechanism of inactivation in marsupials, it seems that the inactivation of the paternal chromosome was the ancestral condition. The functional quasi-heterozygosity for X-linked genes in the eutherian female may have been the selective force that led to the evolution of random inactivation in this mammalian subclass.

Whether or not imprinted inactivation is the ancestral state of X-chromosome silencing, it is not clear why the paternal chromosome is exclusively affected. From an evolutionary point of view, paternal inactivation may be the result of the genetic conflict of interest between the maternal and paternal genes within the embryo (Hurst, 1997). A mechanistic explanation for paternal X-chromosome inactivation in the extraembryonic tissues of eutherians may be found in the fact that the *XIST/Xist* is expressed and the X chromosome silenced during male meiosis (Heard et al., 1997). Since the trophoblast is the first to develop and undergoes X inactivation before the epiblast, perhaps there is insufficient time to erase the inactivating epigenetic pattern acquired during spermatogenesis and the paternal X is silenced by default. This may be especially significant in rapidly developing animals such as the mouse and rat, which also show paternal inactivation in the extraembryonic tissues (Wake et al., 1976). The human embryo and

supporting tissues may simply have more time to develop and therefore undergo random inactivation. Unfortunately, there are no data concerning the parental origin of the extraembryonically inactivated chromosome in other eutherian mammals, which may shed light on the problem. Further support for this theory comes from studies on the appearance of macroH2A and its association with the inactive X to form the macrochromatin body (MCB). In imprinted inactivation in the trophoblast the MCB is formed at a much earlier stage than in ES cells, perhaps due to its association with the sex vesicle formed by the X and Y chromosome in the eutherian male meioses (Solari, 1993, Avner and Heard, 2001).

Interesting evidence from cloned mouse embryos seems to support this theory (Eggan et al., 2000). Mice cloned from adult somatic cells exhibit random inactivation in the embryo proper, indicating that the epigenetic mark of somatic cell inactivation has been erased during the cloning process. However, the silencing pattern in the extra-embryonic cells is not random. The X chromosome in the female somatic cells that was induced to be inactive through cell culture techniques will be inactive in the trophoblast, possibly due to its inactive chromatin configuration at the time of transfer. In contrast, mice cloned from ES cells, in which the two X chromosomes are equivalent and theoretically lack the epigenetic mark, since they have not yet undergone silencing, will have a random inactivation pattern in both the trophoblast and epiblast tissues. These results point to the existence of a gametic mark on the paternal X chromosome, possibly in the promoter of *Tsix* (Lee, 2000), which predisposes it to inactivation and may be a result of its silenced status during gametogenesis. In the absence of the somatic or embryonic epigenetic mark, inactivation occurs randomly.

Beyond the human and mouse models of X-chromosome inactivation

It has now been well established that the *XIST/Xist* gene is essential for X-chromosome inactivation in mouse and humans, in contrast to the situation in marsupials where its very existence is in doubt. Considerably less research has been done on other eutherian mammals, but the few reports in cows, voles and rats seem to conform with the mouse/human model. The *Xist* RNA has been seen to coat the vole and rat inactive X chromosome (Duthie et al., 1999). The only non-rodent organism studied has been the cow and recent evidence from bovine embryos shows *Xist* expression in female embryos (De La Fuente et al., 1999). Apart from the 5' fragments of lepine, equine, primate and bovine *Xist* sequenced for the evolutionary conservation studies of Hendrich et al., (1993), and a few small and scattered bovine, feline, porcine and rat ESTs, the only full *XIST/Xist* sequence is for the human, mouse and three voles (Brockdorff et al., 1992, Brown et al., 1992, Nesterova et al., 2001 *in press*). With the exception of the gene sequences, which have been deposited in GenBank, no other information regarding vole *Xist* is available to the public at this time (Nesterova et al., 2001 *in press*). The identification of conserved regions of the gene through mouse – human comparisons have not been helpful in the search for functional domains nor a marsupial homologue, indicating that it is time to look beyond the traditional human and mouse models. Although, *Xist* sequences have been recovered from all the eutherians examined so far, no studies have been done in the Insectivora, an order that is traditionally thought to represent an early eutherian lineage (Krettek et al., 1995). The only indication that X-inactivation occurs in this order comes from a study by

Jegalian and Page (1998), in which differential methylation was seen for three out four X-linked genes studied in the male and female European hedgehog.

While the evolutionary relationships between the mammalian subclasses are well established, the phylogeny within the Eutheria is still a subject of considerable debate (Novacek, 1992, Cooper and Fortey, 1998, Cao et al., 2000, Madsen et al., 2001, Murphy et al., 2001). The eutherians are thought to have had two periods of rapid radiation, which have resulted in a star-like or bush-like phylogeny, which effectively means that all of the lineages are more or less equidistant from each other (Cao et al., 2000). However, some authors have challenged this star-shaped phylogeny and maintain that there are deeper branches and divisions than was previously thought (Cooper and Fortey, 1998). The oldest radiation occurred around 102 MYA which coincides with the separation of the African and South American continents (Cao et al., 2000). The second diversification at the Cretaceous-Tertiary boundary (65 MYA) is thought to be the result of the demise of the dinosaurs and the availability of new environmental niches for the extant mammals (Cao et al., 2000). Despite the intense interest in the field, the eutherian phylogeny, especially at the deeper levels, still remains unresolved (Cooper and Fortey, 1998, Cao et al., 2000).

An especially contentious problem has been the monophyly and position of the order Insectivora (Liptophyla) within the higher level relationships of the placental mammals. On the basis of morphology the insectivores occupied a basal position in traditional eutherian phylogenies and were thought to represent the closest living relatives to the ancestral placental stock (Novacek, 1992). According to this morphological view, which was the prevalent model until recently, the insectivores were a monophyletic group consisting of the Soricidae (shrews), Talpidae (moles), Erinaceidae (hedgehogs and gymnures), Solenodontidae (solenodons),

Tenrecidae (tenrecs) and Chrysochloridae (golden moles) (Stanhope et al., 1998). Three independent studies based on nuclear and mitochondrial sequence have effectively shattered the monophyly of the Insectivora (Stanhope et al., 1998, Madsen et al., 2001, Murphy et al., 2001). All of the studies have consistently placed the African insectivores, the tenrecs and golden moles, within the new order Afrotheria. Despite these significant results, the position and grouping of the remaining insectivores (shrews, moles, solenodons and hedgehogs), which now seem to be called the Insectivora, Liptophyla or Euliptophyla, still remains uncertain.

Analysis of the complete mitochondrial DNA of the European hedgehog placed the insectivores basal to all the other eutherian sequences (Krettek et al., 1995). However, Mouchaty et al. (2000) rejected a monophyletic relationship between the hedgehog and the mole, based on the analysis of complete mitochondrial genome of the European mole. They showed the mole and shrew to be a sister group to the Chiroptera and proposed the dissolution of the Order Liptophyla. This group was supported by another study of complete mitochondrial genomes by Cao et al. (2000).

On the other hand, the extensive studies of Murphy et al. (2001), which used 10,000 base pairs of nuclear and mitochondrial sequence, support the monophyly of the shrews, moles and hedgehogs (the Euliptophyla) and place the group in a superordinal clade (Laurasiatheria) along with the Cetariodactyla (even-toed animals and whales), Perissodactyla (cloven-toed animals), Carnivora, Chiroptera (bats) and Pholidota (pangolin). Laurasiatheria, meaning "from the area of Laurasia" or Europe, Asia and North America was suggested for this new clade, although no statistical support was provided. Similarly, Madsen and colleagues (2001) place the hedgehog, mole and shrew together and position them within a comparable superordinal group, based on approximately 8000 base pairs of nuclear and mitochondrial sequence. The results of Stanhope

et al. (1998) derived from several combinations of mitochondrial and nuclear genes also support the Liptophylan clade of shrews, moles and hedgehogs. Interestingly, the results of both Madsen et al. (2001) and Murphy et al. (2001) place the primates and rodents into the same superordinal clade, which never includes the core insectivores. Additionally, both studies find the core insectivores and the African insectivores (tenrecs, golden moles) at the base of each of their respective superordinal clades, an observation that is consistent with the traditional view that insectivore-like forms gave rise to the various eutherian taxa. In a very recent, and probably the largest, phylogenetic study concerning the eutherian radiation the authors used a novel approach and combined all the known molecular and morphological phylogenies into several parsimonius "supertrees" (Liu et al., 2001). In all three "supertrees" presented, the Soricomorpha and Erinaceomorpha formed a monophyletic group that was distinct and distant from both primates and rodents.

A review of all the available phylogenetic studies concerning the core insectivores (Erinaceidea, Talpidae, Soricidae and Solenodontidae) reveals that the grouping is rejected by two analyses of complete mitochondrial genomes (Cao et al., 2000, Mouchaty et al., 2000), yet it is supported by morphological (Novacek, 1992) and nuclear data and combined nuclear and mitochondrial data (Liu et al., 2001, Madsen et al., 2001 Murphy et al., 2001). The higher rate of substitution (more informative sites) and smaller effective population size (shorter coalescence times) in comparison to nuclear protein-coding genes, has made mitochondrial DNA the genome of choice for studying closely related species (Springer et al., 2001). However, a recent paper by Springer and colleagues (2001) showed that nuclear exons are more efficient and consistently achieve greater resolving power on a per-residue basis in comparison with mitochondrial protein-coding or rRNA genes. In an attempt to recreate several well-

established eutherian phylogenies using various sets of nuclear and mitochondrial data sets it was found that the nuclear genes were more successful in recovering the benchmark clades. Although mitochondrial DNA sequence are invaluable in resolving more recent divergences, it seems that nuclear genes should be used when recovering deep mammalian divergences, especially if the mitochondrial data does not agree with the morphological evidence. In conclusion, the authors argue that the most valuable data set would be one composed of both nuclear and mitochondrial sequences, since they provide two independent estimates of the phylogenetic relationship. In light of this study, the monophyly of the core insectivores, which is supported by all the current phylogenetic studies, except those relying exclusively on mitochondrial data, may prove to be correct.

Although the phylogenetic position and monophyly of the Insectivora has not been unequivocally established, molecular evidence seems to support a basal position for the group, at least within the Laurasiatheria. More importantly for the purposes of this study, none of the analyses show a close phylogenetic relationship between the Euliptophylans and primates or rodents, indicating that *Xist* sequence from a member of this order may provide new information regarding the evolution of the gene. The identification of the X-chromosome inactivation process and the *Xist* gene in an insectivore would indicate that *Xist*-mediated silencing evolved before the eutherian radiation and is shared by all members of this subclass. Furthermore, insectivore *Xist* sequence may provide insight into the function of the gene and may indicate areas of conservation that may facilitate the identification of the elusive marsupial homologue.

Chapter 2

Materials and Methods

Mole samples and cell lines

Prior to the establishment of Coast mole cell lines I had been using mole DNA kindly given to me by Dr. Kevin Campbell (at the Department of Zoology at the University of British Columbia) and had some frozen organs from several moles found dead on the road or in mole traps. The RNA extracted from these samples was either nonexistent or of very poor quality. Fresh mole samples were finally obtained from three female moles sacrificed in a study by Dr. Campbell. Additional animals were acquired in the spring of 2000 during a short mole hunting "field season" – an endeavor which would have been impossible without the help of Dr. Campbell and Tim Sheehan. RNA was extracted from these new samples, which were also used to create Coast mole cell lines that served as additional source of RNA. Unfortunately, most of these moles were females and only one male animal, found in a trap, was sufficiently preserved to allow the establishment of a cell line. Information on the identity and history of all the mole samples is presented in Table 1.

The mole cell lines, all of which were primary cultures, were prepared from fresh kidneys, which were placed in Minimal Essential Medium (α MEM) (Gibco/BRL) immediately after collection. The organs were then rinsed with 70 % ethanol in the tissue culture hood to reduce bacterial contamination, followed by a rinse in α MEM. The tissue was then diced on a

Coast mole	Source	Date	Cell Line	Comments
CM1 (female)	Dr. K. Campbell	May 1998	-	Received DNA only
CM3 (male)	Dr. K. Campbell	May 1999	-	Found dead on a road
CM4 (male)	Dr. K. Campbell	Feb. 2000	-	
CM5 (male)	Dr. K. Campbell	May 1999	-	
CM6 (male)	Dr. K. Campbell	May 1999	-	
CM7 (female)	Dr. K. Campbell	May 1999	-	
CM8 (male)	Captured *	March 2000	-	Accidentally killed in the field
CM9 (male)	Captured *	March 2000	-	Accidentally killed in the field
CM10 (female)	Captured *	March 2000	-	Drowned
CM11 (female)	Captured *	March 2000	+	Sacrificed
CM12 (female)	Captured *	March 2000	+	Sacrificed
CM13 (female)	Captured *	March 2000	+	Sacrificed
CM14 (male)	Tim Sheehan	May 2000	-	Found in trap (dead <24 hours)
CM15 (male)	Tim Sheehan	May 2000	+	Found in trap (dead <1 hour)
CM16 (female)	Tim Sheehan	June 2000	+	Found in trap
CM17 (female)	Tim Sheehan	June 2000	+	Found in trap

* Captured in the field with Dr. K. Campbell and T.Sheehan

Table 1: Coast mole samples

coverslip using a sterile razor blade and transferred to a fresh tube with 1000units/mg collagenase (Sigma), an enzyme, which destroys the tissue matrix, and releases the individual cells. After 20 minutes of incubation at 37° C, the sample was briefly centrifuged (3500 rpm). Once the supernatant had been removed, the cell pellet was rinsed in α MEM and centrifuged once again. The cell pellet was then incubated with 0.1 ml of 0.25 % trypsin-EDTA (Gibco/BRL) at 37° C for 5 minutes. Following this incubation, the cells were centrifuged, rinsed with α MEM and resuspended in 200-500 µl of α MEM supplemented with 7.5 % fetal calf serum (Cansera), 1 % MEM nonessential amino acids, 1% L-glutamine and 1% penicillin and streptomycin (all from Gibco/BRL). Before the cells were plated trituration was used to break up any remaining cell clumps. The cells were incubated at 37° C, 5 % CO_2 and 90 % humidity.

To obtain cell pellets for the extraction of genomic DNA, the medium was removed, the cells were washed with phosphate saline buffer (PBS), trypsinized (0.25 % trypsin-EDTA) and resuspended in fibroblast medium. The Coast mole cells proved very hardy and had to be scraped off the plates after trypsinization. The cell suspension is then centrifuged and the supernatant removed leaving the cell pellet. The protocol for freezing cell lines was very similar to that for producing cell pellets, except that the cells were resuspended in fibroblast medium enriched with 15 % fetal calf serum and 10 % DMSO (dimethyl sulfoxide) after trypsinization. The cell suspension was then placed in an isopropanol filled container at –70 °C and allowed to cool slowly, after which they are transferred to liquid nitrogen for long term storage.

Extraction of genomic DNA

In order to extract genomic DNA, cell pellets or homogenized tissue were placed in Tris-EDTA (TE) buffer (pH 7.5-8) and incubated over night at room temperature with $1/20^{th}$ volume of 20 % SDS (sodium dodecyl sulfate) and a pipette tip of proteinase K – a proteolytic enzyme active against a broad spectrum of proteins. These solutions serve to rupture the cell membranes and digest proteins, thus releasing the nucleic acids. The following day $1/10^{th}$ volume 5 M NaCl was added, the tube was shaken vigorously and spun for 15 minutes (2500 rpm). The supernatant was then removed from the pellet of denatured proteins. Additional salt (5 M NaCl, one third of the original volume) and SDS ($1/30^{th}$ of original volume) were added, the preparation was again shaken vigorously and spun at 2500rpm for 15 minutes. The supernatant was then transferred to a clean tube and two volumes of 95 % ethanol were added to precipitate the DNA, which was carefully spooled, resuspended in double distilled water (ddH2O) and stored at 4° C.

RNA isolation

Total RNA was isolated using the acid-guanadinium-phenol-chlorofom RNA protocol, also known as the RNAzol B method, developed by Chomczynski and Sacchi (1987). The guanadinium and water molecules form complexes with the RNA, abolishing the hydrophilic interactions with the DNA and proteins, thus the DNA and proteins are removed from the aqueous phase, in which the RNA remains. Phenol acts as a denaturing and deproteinating

agent, while chloroform removed any traces of phenol from the preparation. The first step in the protocol involved the homogenization and vortexing of tissue in Solution D (4 M guanadinium thiocyanate, 25 mM sodium citrate pH7, 0.5 % sarcosyl, 0.1 M 2-mercaptoethanol (added just before use). Normally 0.6 ml of Solution D were used per 100 mg tissue or a confluent 60 mm tissue culture dish. Homogenization was not necessary for tissue culture cells since vortexing is enough to disrupt the cell membranes. After the addition of equal volume of diethyl pyrocarbonate-treated (DEPC) ddH2O saturated phenol and $1/10^{th}$ volume 2 M sodium acetate (pH4), the sample was mixed by inversion and vortexing. Chloroform (0.4 volume) was added to the solution and the mixture was vortexed and placed on ice for 5 to 15 minutes. After centrifugation (10 minutes, 13500 rpm) the RNA-containing supernatant was removed, while the proteins and DNA remained in the lower organic layers. After the addition of an equal volume of isopropanol, the solution was left at –20° C overnight. A 10 minute centrifugation at 13500 rpm precipitated the RNA, which could then be washed with 70% ethanol and reprecipitated (10 min, 13500rpm). The air-dried RNA pellet was resuspended in 10-100 µl of DEPC-ddH20 and stored at -20° C.

DNAse treatment of RNA

RNA in DEPC-ddH2O was combined with porcine RNAsin ($1/20^{th}$ total volume, Amersham Pharmacia Biotech) and RNAse-free DNAse ($1/10^{th}$ total volume, Epicentre technologies) and incubated at 37° C for one hour. The volume of the solution was then increased to 200 µl with DEPC-ddH2O and an equal volume of 1:1 phenol-chloroform was

added to denature any proteins. The sample was then vortexed, placed on ice and centrifuged for 5-10 minutes at 12500 rpm. The aqueous upper layer was transferred to a new tube where an equal volume of chloroform was added to remove any traces of phenol. The sample was vortexed and centrifuged once again. The aqueous phase was transferred to a new tube to which 0.15 volume of 2M sodium acetate in DEPC-ddH2O and an equal volume of isopropanol were added to precipitate the DNA (-20° C, overnight). The following day, after centrifuging, the supernatant is discarded and the RNA pellet was resuspended in DEPC-ddH2O.

The Polymerase chain reaction (PCR)

Approximately 100 ng of template were added to a 25 µl PCR reaction containing 20 µM each dNTP, 1.5 mM $MgCl_2$, 1 µl 10X PCR buffer, 1 µM primer and 0.625 units *Taq* DNA polymerase (all reagents , with the exception of the primers, are from Gibco/BRL). The reactions were overlayed with 20 µl mineral oil and forty cycles of 1 minute at 94° C, 1 minute at 50° C and 2 minutes at 72° C were performed in a Techne Genius Thermal Cycler. Unless otherwise specified, these conditions were used in all the PCRs. The amplification of the CpG islands in the methylation assay required a change in the $MgCl_2$ concentration and the addition of betaine (Sigma), which improves the PCR amplification of GC-rich sequences by reducing the formation of secondary structure (Henke et al., 1997).

Long PCR, using the ExpandLong Template system (Boehringer Mannheim), was performed on fragments greater than 1 kilobase. Two 25 µl reaction mixes, one containing the enzyme mix (2.5 units), 5ul 5x Buffer 3 (0.75 mM $MgCl_2$) and ddH20, and the other composed

of 0.5 mM of each dNTP, 0.3 μM of each primer and approximately 200 ng of genomic DNA, were combined and overlayed with 30 μl mineral oil. The mix was incubated at 94° C for two minutes and then cycled forty times (30 seconds at 94° C, 30 seconds at a 50-54° C, elongation at 68° C). The elongation times were varied according to the manufacturer's recommendations and ranged from three to eight minutes. A final elongation step was performed at 68° C for seven minutes.

Inverse PCR

Inverse PCR is a useful method for obtaining additional sequence flanking a known fragment. In this technique, genomic DNA is digested with a restriction enzyme that does not cut within the known area and then ligated. Primers are designed from the known sequence, but pointing out of the region, and are used for PCR amplification of the ligated DNA (see Figure 3). Coast mole 10 genomic DNA (1 μg) was digested with the appropriate enzymes overnight in a 10 or 20 μl reaction. Restriction enzymes with a four base pair recognition site, such as *AluI*, are particularly useful since they produce relatively small fragments that can be PCR amplified. The following day the digest was combined with 4000 units of T4 DNA ligase (NEB), 50 μl 10X ligase buffer and ddH20 was added increase the volume to 500 μl. The ligation reaction was left at room temperature overnight. The DNA was precipitated with the addition of salt and two volumes of 95 % ethanol, resuspended in 10-20 μl of ddH2O and 1-2 μl of this suspension were used in a PCR reaction (see above for details on ethanol precipitation).

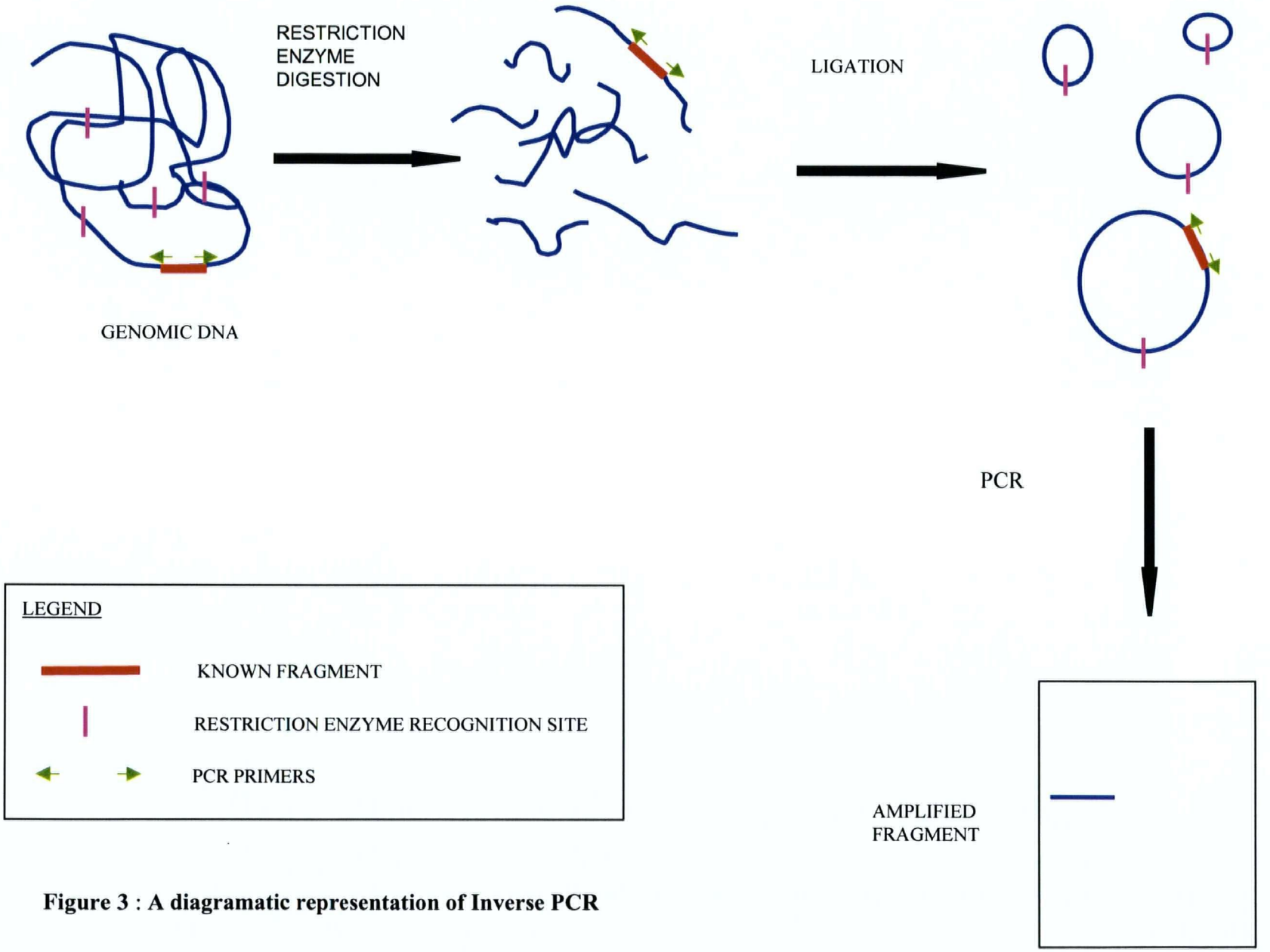


Figure 3 : A diagramatic representation of Inverse PCR

All the PCR products were run on 1-2.5 % agarose gels, stained with ethidium bromide and visualized under UV light. PCR primers were selected manually and the oligonucleotides were made at the NAPS facility at the University of British Columbia.

The primer sequences, sources and amplification conditions for the primers designed from conserved *XIST/Xist* sequence, mole-specific primers and non-*XIST/Xist* primers are presented in Figure 4 and Table 5 (Chapter 4), and Tables 2 and 3, respectively.

Reverse transcription

Five nanograms of RNA in DEPC-ddH2O were mixed with 5X first strand buffer (1/5th volume, Gibco/BRL), 0.1 M DTT (dithiothreitol, $1/10^{th}$ volume), 1.25 mM deoxynucleotide triphosphates (dNTPS, $1/20^{th}$ volume, Gibco/BRL), $1/20^{th}$ volume random hexamer primers and $1/20^{th}$ volume M-MLV (Moloney Murine Leukemia virus) reverse transcriptase (MLV-RT, Gibco/BRL) and incubated at 42° C for two hours. The RT enzyme uses the presence of RNA to synthesize a complementary DNA strand, which can serve as a template in PCR. Although the principle is the same, the SuperScript (Gibco/BRL) reverse transcriptase requires slightly different conditions, which are described in detail in the Gibco/BRL product manual (Cat No.18064-014).

```
COW(AF10496)     TTGCTGCAGGGACAATATGGC
HORSE(U50911)    TTGCCGCAGGGACAATATGGC
RABBIT(U50910)   TTGCCGCAGGGACAATATGGC
HUMAN(M97168)    TTGCCGCAGGGACAATATGGC

SKXIST1          TTGC(T/C)GCAGGGACAATATGGC
```

```
COW(AF10496)     ATCCGACCCCAGCATTAGCC
HORSE(U50911)    ATCCGACCCCAGCATTAGCC
RABBIT(U50910)   ATCCGACCCCAGCATTAGCC
HUMAN(M97168)    ATCCGACCCCAGCATTAGCC

SKXIST2          ATCCGACCCCAGCATTAGCC
```

```
COW(AF10496)     GCCACTTGCAGTGCTGGATA
HORSE(U50911)    GCCTCTTGCAGTGCTGGATA
RABBIT(U50910)   GCCTCTTGCAGTGCTGGATA
HUMAN(M97168)    GCCTCTTGCAGTGCTGGATA

SKXIST3          GCC(A/T)CTTGCAGTGCTGG
```

```
COW(AF10496)     CCTAATATCCAGCACTGCAAG
HORSE(U50911)    CCAGCTATCCAGCACTGCAAG
RABBIT(U50910)   CCAGATATCCAGCACTGCAAG
HUMAN(M97168)    CCAGATATCCAGCACTGCAAG

SKXIST4          CC(T/A)(A/G)ATATCCAGCACTGCAAG
```

```
COW(BE483406)    TAATAGCTACAAAGTAGTACTTAC
HUMAN(M97168)    TAATAGCTACGAAGTAGTACTTAC

SKXIST6          TAATAGCTAC(G/A)AAGTAGTACTTAC
```

```
COW(BE483406)    ACCACACTCAAGTGAGGACTT
HUMAN(M97168)    ACCACACTGAGGTGAGGACTT

SKXIST7          ACCACACT(G/C)A(G/A)GTGAGGACTT
```

```
COW(AV617038)    TTTNACGATTCCTAGGTGGA
HUMAN(M97168)    TTTGACGATCCCTAGGTGGA
MOUSE(L04961)    TTTGACGATCCCTAGGTGGA

SKXIST9          TTTGACGAT(T/C)CCTAGGTGGA
```

```
COW(AV617038)    TAGCATTGCAGG-CATGTTGA
HUMAN(M97168)    TAGAGTGCCAGG-CATGTTGA
MOUSE(L04961)    TAGGGTAGCAGTGCATCTTGA

SKXIST10         TAGN(G/A)TN(G/C)CAGGCATGTTGA
```

```
CAT(AF197966)    ACAAGAG-TAC-CAGATGTGCC
HUMAN(M97168)    ACAAGAAATACACAGATGTGCC

SKXIST11         ACAAGA(A/G)ATACACAGATGTGCC
```

```
CAT(AF197966)    CCTTGGGGCCTCACTCTGTC
HUMAN(M97168)    CCTTGGGACCTCGCTTTGTC

SKXIST12         CCTTGGG(G/A)CCTC(A/G)CT(C/T)TGTC
```

```
HORSE(U50911)    TCACTTCTTAAAGCGCTGCA
RABBIT(U50910)   TCAATTCTTAAAGCGCTGCA
HUMAN(M97168)    TCAGTTCTTAAAGCGCTGCA

SKXIST13         TCANTTCTTAAAGCGCTGCA
```

Figure 4: Conserved *Xist* primers.
The locations of the primers are shown in Figures 8,12,13 and 15. The exact locations and amplification conditions of the primers are presented in Table 5 (Chapter 4).

Primer	Sequence	Conditions	Comments
CMXIST1	cattgctgaagtggcctgagg	50-54° C, 40X	Positive for expression with CMXIST2
CMXISTREV1	cctcaggccacttcagcaatg	50-54° C, 40X	
CMXIST2	gttcctcttgagaaaggcaagg	50-54° C, 40X	Positive for expression with CMXIST1
CMXISTREV2	ccttgcctttctcaagaggaac	50-54° C, 40X	
CMXISTREV3	aaggccaattaatgagttca	50-54° C, 40X	
CMXISTREV4	tccatactgccttatcactg	50-54° C, 40X	
CMXISTREV5	atcaggcaacaactcactgc	50-54° C, 40X	Positive for expression with CMXIST10
CMXISTREV6	gaacagcagttctttgtaatc	50-54° C, 40X	
CMXISTREV7	cctcaattccgtgactgtag	50-54° C, 40X	
CMXISTREV8	gatactagagtaactgcagcg	50-54° C, 40X	
CMXIST10	caggtggagttgataacctgg	50-54° C, 40X	Positive for expression with CMXISTREV5
CMXIST11	agcactgctcagaagcaatgc	50-54° C, 40X	

Table 2: Primers designed from Coast mole *Xist* sequence.
The locations of the primers are indicated in Figures 9, 11, 12 and 13. The "REVERSE" primers are referred to in the text as CMXISTREVERSE. For example, CMXIST1REVERSE is abbreviated to CMXISTREV1.

Primer	Sequence	Mole amplification	Conditions	Source	Comments
mXist1 mXist2	cacatcttctccacttgaga gccactactatgagcagga	NO		C. Brown	Amplify human and mouse *Xist*; position in human *XIST* (13503-13522 and 13925-13906)
SKHPRT1 SKHPRT2	ggaaagaatgtcttgattgttga ggaatttcgaatccaacaaagtc	YES (cDNA only) ~150 bp band	52° C, 40X	This work	Designed from sequenced conserved between human(gi4504482), dog (gi786012), pig(gi4809240), rabbit(gi2655050), golden hamster (gi2896165), platypus(gi4322123), wallaby (gi4322117) and opossum (gi432121) genes.
ZFX1 ZFX2	gagctcggagctgacaaaaa ctacccttccgcattttcct	YES ~105 bp band	1M betaine 50° C, 40X	Luoh et al., 1995	Derived from CpG island conserved between mouse and human
SMCX1 SMCX2	cctcgggcccaccatggcg ctgattttcgcgatgtagcc	YES ~117 bp band	0.5mM$MgCl_2$, 60° C, 40X	Jegalian and Page, 1998	Derived from CpG island conserved between mouse and human
ALD1 ALD2	gtgacatgccggtgctctcca gcgtgcaggaatacccggttcat	NO		Jegalian and Page, 1998	Expected product is 226bp, but under low stringency conditions the only band in the mole is 100bp, which does not contain methylation sensitive site. Derived from CpG island conserved between mouse and human
FRAX1 (7) FRAX2 (2)	gctcagctccgtttcggtttcacttccggt agcccccgacttccaccaccagctcctcca	YES ~250 bp band	2mM $MgCl_2$, 2M betaine, 50° C, 40X	Fu et al., 1991	CpG island. Primers flank polymorphic CGG repeat.
ARA1 (A) ARA2 (B)	tccagaatctgttccagagcgtgc gctgtgaaggttgctgttcctcat	YES ~250 bp band	1M betaine, 52° C, 40X	Cutler –Allen et al., 1992	CpG island. Primers flank polymorphic CAG repeat.
TeSRY1 TeSRY2	cgctgcctgtagtctctgtg tcatggtttggtctcgagatc	YES ~158 bp band	50° C, 40X	Sanchez et al., 1996b	Male specific

Table 3. Miscellaneous primers

With the exception of the mXist1,2 pair, none of the primers in this figure are located in the *Xist* gene. The ALD1,2, FMR1,2, ARA1,2 and ZFX1,2 pairs were used in the methylation analysis described in Chapter 2, the TeSRY1,2 pair was used for the sexing of animals and the SKHPRT1,2 primers served as a positive cDNA control.

Methylation-sensitive restriction digests

A reaction mix containing 20-200 ng of genomic DNA, 5X DBS (5 mM DTT, 0.5 mg/ml bovine serum albumin and 10 mM spermidine), 1 µl 10X of the appropriate buffer and the restriction enzyme (10 Units) was incubated at 37° C overnight. The enzyme was heat-inactivated in a 10 minute incubation at 65° C the following day. One microliter of the digest was used in PCR amplifications using primers flanking CpG sites and the products were visualized under UV on an agarose gel stained with ethidium-bromide. The methylation sensitive enzyme *HpaII* (NEB) was used to evaluate the methylation status of the CpG sites, while *HindIII* (Gibco/BRL) served as a negative cutting control.

DNA purification for sequencing and Xist sequence analysis

Gel extraction and purification of PCR fragments was performed using the QIAquick PCR purification kit (QIAGEN Cat. No.28704), according to the enclosed protocol (page 24). Briefly, the PCR band was excised from an agarose gel stained in ethidium bromide and combined with a solubilization buffer (Buffer QG). The liquefied agarose was then placed in a silica-gel membrane spin column, to which the DNA should adsorb, and centrifuged. Additional buffer QG was added to the column and centrifuged to remove all traces of agarose. The column was washed several times in an ethanol based buffer (buffer PE) to remove any salts and the DNA was eventually eluted in 30 to 100 µl ddH20. The purified fragment was sent to the NAPS laboratory at UBC for sequencing.

Homology searches were performed using the BLAST programs (Altschul et al., 1990), while the sequence alignments were done in CLUSTALW v. 1.6 set to DNA and using IUB matrix (Thompson et al., 1994) and available at the Canadian Bioinformatics Resource site (*www.cbr.nrc.ca*). The alignments were not manually improved. The ORF-Finder (www.ncbi.nlm.nih.gov/gorf/gorf.html) and GRAIL programmes (Uberbacher et al., 1991) were used to search for open reading frames.

Chapter 3

X-chromosome inactivation in *Scapanus orarius*

Introduction

The Coast mole (Scapanus orarius)

The focus of this thesis is the study of X-chromosome inactivation and the *Xist* gene in the Coast mole, *Scapanus orarius* (Mammalia, Insectivora, Talpidae, Talpinae) (Figure 5). Before any attempts at identifying an *Xist* homologue were made and in order to evaluate any future results, it was necessary to determine whether these animals underwent X-chromosome inactivation. The detection of X-chromosome silencing in *Scapanus* is the topic of this chapter.

The Coast mole (Figure 5) is a representative of the family Talpidae whose members are distributed widely throughout North America and Eurasia and include moles, shrew-moles and desmans (Yates and Moore, 1990). The unique pectoral anatomy necessary for the burrowing lifestyle and other traits have provided morphological evidence for the monophyly of the talpid group whose members enjoy fossorial, terrestrial and aquatic lifestyles. The Coast mole, also known as the Pacific mole, belongs to an exclusively terrestrial genus and is found from the Fraser Delta in British Columbia in the north to the coastal areas of northwest California in the south (Nagorsen, 1996). Like all fossorial moles, the Coast moles spend most of their lives underground, preferably in well-drained and soft soil, and are solitary except during the mating season (Yates and Moore, 1990).

Figure 5: The Coast mole, *Scapanus orarius* (courtesy of S. Tim Sheehan).

Although superficially similar to the only other underground mammals, the fossorial rodents, the talpids are a much older and more widely distributed lineage. The family was widespread and diverse during the Oligocene, while the oldest fossils date to the late Eocene (approx. 36 MYA) in Europe and the Miocene (approx. 24 MYA) in North America (Yates and Moore, 1990). Additionally, unlike the subterranean rodents, which are well known to exhibit high levels of chromosomal variation and polymorphisms, the talpids have a remarkable degree of chromosomal conservation despite their antiquity. The vast majority of examined species, including *S. orarius*, have a diploid number of 34 including an XX or XY pair (unpublished data in Yates and Moore, 1990). The Y chromosome is minute (Yates and Moore, 1990) and synapses with the X chromosome during spermatogenesis to form the XY body (Solari, 1993).

Sex determination mechanisms in the mole, which have only been extensively studied in the European mole (*Talpa europea*) and other members of the genus *Talpa*, conform to the mammalian XX, XY system. The reason for the intense interest in sex determination of the European talpids was due to a large number of apparently hermaphroditic XX individuals. Initial studies indicated that these animals had functional ovotestes (Jimenez et al., 1993, Sanchez et al., 1996a), but these results were later challenged after no evidence of spermatogenesis was found (Whitworth et al., 1999). Further experiments showed that all the male European moles carry the *Sry* gene (Sanchez et al., 1996b) and that the "hermaphrodites" were chromosomally normal females. Although the testicular tissue of the ovotestis does not produce viable gametes, the presence of bilateral ovotestes in all the females of this species makes them unique among mammals. Additionally, the ovarian and testicular components of the female ovotestis seem to fluctuate in size seasonally. The ovarian portion is increased during mating season, while the testicular region, which produces the androgens responsible for the

observed masculinization, is at its largest during the autumn when the increased aggressiveness and territoriality of the female may carry significant evolutionary benefit. Sanchez et al. (1996b) obtained the *Sry* DNA-binding HMG box sequence from other mole species, as well as from the Algerian hedgehog and several shrew species, and found that the insectivore sequences bear more resemblance to the human than to the mouse homologue. This analysis also confirmed the long-standing view that the moles and shrews are very closely related (Suborder Soricomorpha), while the hedgehog may represent a more distant group (Suborder Erinaceomorpha) within the Euliptophyla (Sanchez et al., 1996b).

Although the karyotypes of several mole species have been published (Yates and Moore, 1990), there have been no reports regarding the presence of a Barr body or any other indications of X-chromosome inactivation. The absence of the Barr body is not surprising since it is not seen in many mammals, especially those with large blocks of heterochromatin at other locations of the genome, and is not a very reliable indicator of X silencing (Solari, 1993). In the absence of a Barr body, X-chromosome inactivation can be established through molecular methods such as assaying the expression of various X-linked genes in cell culture, determining the methylation status of genes on the X, observing replication timing differences between the two X chromosomes or performing fluorescence *in situ* hybridization (FISH) for *Xist* or other X-linked RNAs. To establish the existence of X inactivation in a somatic cell culture it is necessary to obtain sequence polymorphisms to distinguish between the two alleles present in the genomic DNA and determine which one is expressed. In the absence of a polymorphism, one can use somatic cell hybrids in which the X chromosome of the species of interest is retained in the chromosomal background of the host species. Gene expression can then be assayed through RT-PCR (see Materials and Methods) to determine whether the retained

chromosome is active or inactive. To perform FISH studies, gene sequence has to be known in order to design the nucleic acid probes. Since neither mole somatic-cell hybrids, polymorphic markers nor X-linked gene sequence were available, we decided to investigate the existence of X-chromosome inactivation in the Coast mole by examining the methylation status of X-linked genes in the male and female animals.

DNA methylation in the mammalian genome

DNA methylation is an epigenetic modification that only occurs at the 5' cytosine in a CpG dinucleotide in the mammalian genome in which approximately 70 % of all CpGs are methylated (for review see Goto and Monk, 1998, Robertson and Jones, 2000, Robertson and Wolffe, 2000). The CpG dinucleotide is significantly under-represented in the genome due to the increased spontaneous deamination rate of the 5-methyl-cytosine to thymine, a naturally occurring DNA base, and the inefficient repair of this deamination product. However, the resulting CpG depletion is not randomly distributed and there are regions of the genome in which the dinucleotide can be found at the expected frequency. These regions (known as CpG islands) occur at the 5' end of genes and, although they represent only 15 % of the total genomic CpG sites, account for 50 % of the unmethylated CpG dinucleotides. The approximately 45,000 CpG islands identified are normally unmethylated and are accompanied by an open chromatin structure and gene expression. Apart from the imprinted genes, the only exceptions to this rule are the CpG islands of inactivated genes on the silenced X chromosome, which are methylated.

Methylation was not thought to be present in the genomes of the two invertebrate model organisms, *C. elegans* and *D. melanogaster*, however a small number methylated cytosines have recently been identified in the fly (reviewed in Lyko, 2001). Most of the methylation was in the context of CpT and CpA dinucleotides and was restricted to a very early stage in development. Although it is not known whether these methylation patterns have any functional significance, they certainly explain the presence of a DNA methyltransferase (Dnmt2) and methyl-binding protein homologues (dMBD2/3), which had previously baffled scientists.

Methylation is of vital importance in mammals, although the reasons for its existence are not completely understood. One view, that may have been put in jeopardy by the recent discoveries in *Drosophila*, holds that methylation serves to control transcriptional noise and is only necessary in large, complex genomes. Another theory proposes that methylation arose as a silencing defense mechanism against invading DNA sequences. This interpretation is bolstered by the observation that most of the hypermethylated CpG dinucleotides reside within parasitic DNA elements, such as retrotransposons, and lead to their transcriptional silencing. Additionally, kangaroo hybrids show a dramatic expansion of endogenous retroviral-like elements, which is thought to result from the marked hypomethylation of the hybrid genome, lending support to this hypothesis (Waugh O'Neill et al., 1988). Apart from containing the parasitic DNA elements, there is evidence that methylation might stabilize the genome by masking large numbers of homologous repeats and preventing inappropriate recombination through a currently unknown mechanism (reviewed in Roberston and Wolffe, 2000).

Although mouse primordial germ cells and undifferentiated ES cells seem to function without DNA methylation (Lei et al., 1996), several targeted mutations of the genes involved in methylation, the DNA methyltranferases, have shown that methylation is necessary for survival

once cellular differentiation begins (Li et al., 1992, Okano et al., 1999). It is believed that DNA methylation, along with other chromatin modifications, functions to divide the genome into active and inactive compartments, thus enabling the transcription machinery to focus on a limited number of genes necessary for a particular cell's survival (reviewed in Robertson and Wolffe, 2000). Methylation levels and patterns are tissue-specific and vary developmentally. During gametogenesis most parental methylation marks are erased and the gamete DNA is generally hypomethylated, although sperm methylation levels are higher than those of the ovum (reviewed in Goto and Monk, 1998). In preimplantation embryos, there is a further loss of methylation leading to an almost globally demethylated state. However, some methylation marks, such as those on the CpG islands of imprinted genes, seem to be more resistant to demethylation and can persist until the blastocyst stage. Upon implantation and during gastrulation new epigenetic patterns are set down by *de novo* methylation and maintained throughout the lifetime of the organism by the maintenance methyltransferases.

In mammals, DNA methylation is achieved through the action of the DNA methyltransferases (DNMTs), some of which are capable of *de novo* methylation, while others, the maintenance methyltransferases, just duplicate the set methylation patterns. So far, three independent methyltransferases have been identified: DNMT1, DMNT3A and DNMT3B. DNMT1 is the most abundant methyltransferase in somatic cells and is necessary for proper embryonic development, imprinting and X-chromosome inactivation (Li et al., 1992, Beard et al., 1995, Panning and Jaenisch, 1996, Sado et al., 2000). Based on the observations that this enzyme has a great affinity for hemimethylated DNA and that it is associated with replication foci, DNMT1is believed to be the principal maintenance methylase.

The DMNT3 group of methyltransferases (reviewed in Robertson and Wolffe, 2000) show an equal preference for hemi- and unmethylated DNA and are thought to be the *de novo* methylases responsible for the wave of methylation following embryo implantation, as well as for the *de novo* methylation of newly integrated retroviral sequences in mouse ES cells. Mouse knockouts of *Dnmt3b* are lethal, while *Dnmt3*a minus mutants are runted and survive only a few days. New evidence, however, suggest that all three methyltransferases may possess both *de novo* and maintenance capabilities and that the division into *de novo* and maintenance categories may not be valid.

Although the connection between methylation and silencing has been well established, the mechanism is not completely understood (reviewed in Robertson and Wolffe, 2000). Early experiments showed that methylated CpG dinucleotides directly interfered with transcription factor binding sites. However, this mechanism could not account for all the observed silencing, since only a subset of transcription factors contain CpGs in their recognition sequence. The second mechanism is mediated through the methyl-binding protein, MeCP2 (Lewis et al., 1992). Jones et al. (1998) have shown that methylated DNA, through MeCP2, has the ability to recruit histone deacetylases, thus providing a link between methylation and inert chromatin structure (see Chapter 1 for more detail).

Methylation on the inactive X chromosome

The methylation at the CpG islands of inactivated X-linked genes is a well- established feature of the inactive X chromosome in eutherian mammals (Heard et al., 1997) and is frequently used to determine whether a gene is inactivated (Luoh et al., 1995, Carrel and Willard, 1996). The methylation status of a gene can be determined by bisulfite sequencing or by a combination of methylation-sensitive restriction-enzyme digestion followed by PCR (or Southern analysis), which is the method used in this study. Genomic DNA is digested with methylation-sensitive enzymes, such as *HpaII* or *HhaI,* which are able to cleave DNA only if their recognition sequence is not methylated at the CpG dinucleotide. In a subsequent step, the polymerase chain reaction (PCR) is performed using primers that flank the enzyme cut site. Only methylated DNA will remain undigested and will provide a template for the PCR, resulting in a band visualized by agarose gel electrophoresis. Thus, the PCR will only be successful if a gene is methylated and inactive. In the course of this assay it is necessary to perform a "mock" restriction digest using an enzyme that does not cut within the sequence of interest, to ensure that the lack or presence of a PCR fragment is due to the restriction digest and not a PCR artifact.

Of special significance to the study of inactivation in moles is the observation that CpG island methylation correlates perfectly with the transcriptional state of a gene in a broad range of eutherians, including an insectivore representative, the European hedgehog (Jegalian and Page, 1998). Jegalian and Page (1998) studied the inactivation status of four X-linked genes (*SMCX, RPS4X, ALD* and *ZFX*)[2] in large group of eutherians mammals that included one insectivore, the

[2]Upper case notation will be used for the genes in the methylation analysis, with the exception of the murine genes.

hedgehog. They found that silencing through methylation at CpG-islands was conserved in all the species observed. Additionally, they observed that genes escaping X-inactivation, such as *ZFX*, were very reliable indicators of the presence of a Y-linked homologue. The authors concluded that the decay of the Y homologue seems to be the necessary step for a gene to be included in the X dosage compensation pathway, and seems to proceed on a gene-by-gene, rather than regional, basis. Of the four genes studied, *ALD* was inactivated in all species observed, while *SMCX*, *ZFX* and *RPS4X* were silenced in some groups, but not in others. *Zfx* appeared silenced in only in the myomorph rodents (mouse, rat, lemming, hamster), probably due to the restricted testis-limited expression of the Y-linked *Zfy1* and *Zfy2* genes, which led to the inactivation of the X chromosome homologue (Luoh et al., 1995). This is unlike the situation in all other eutherians where the *ZFY* gene is well expressed and *ZFX* escapes inactivation (Jegalian and Page, 1998).

To determine if X-inactivation is present in the Coast mole, I examined the methylation status of *ZFX* (zinc finger on the X), *FMR1* (*FRAX*, fragile X mental retardation) and *ARA* (androgen receptor), which are known to possess CpG islands and are X-linked in humans and rodents (Cutler Allen et al., 1992, Eichler et al., 1995, Luoh et al., 1995, Carrel and Willard, 1996, Heard et al., 1997). Attempts were made to perform this assay using two other X-linked genes, *SMCX* (selected mouse cDNA on X and human homologue) and *ALD1* (adrenoleukodystrophy), but proved unsuccessful. In humans *ARA*, *ALD1*, *FMR1* and *SMCX* are located on the conserved region of the X, while *ZFX* lies on a region of the chromosome that is X-linked in eutherians, but autosomal in marsupials (Pask and Graves, 1998). Although no data exist regarding the chromosomal location of these genes in the Coast mole, according to Ohno's Law, genes present on the non-pseudoautosomal region of the X chromosome in one eutherian,

should be present on the X chromosome of all eutherians (Ohno, 1967). Studies examining the methylation status of 5' CpG islands have determined that *FMR1* and *ARA* are subject to X-inactivation in both rodents and humans (Fu et al., 1991, Norris et al., 1991, Cutler Allen et al., 1992), *ZFX* escapes silencing in every eutherian order studied, except myomorph rodents (Jegalian and Page, 1998). The human and mouse *ZFX/Zfx* CpG islands each contain 19 CpG methylation sites and are almost identical at the sequence levels, indicating a selective pressure for conservation, despite the fact that in one species the gene is silent, while in the other it escapes X-inactivation (Luoh et al., 1995).

The inactivation status of *SMCX* is less clear. In general, species that possess the Y-linked homologue (*SMCY)*, such as primates and rodents, do not inactivate *SMCX* (Jegalian and Page, 1998). In contrast, cows and guinea pigs, which lack *SMCY*, silence *SMCX*. However, rabbits and goats seem to possess *SMCY*, yet still inactivate *SMCX* most probably due to the very low expression of the Y-linked homologue (Jegalian and Page, 1998). As mentioned previously, *Smcx* is thought to escape X-inactivation in rodents, but a more detailed analysis of its expression patterns in the mouse have revealed a more complicated picture. It seems that the murine *Smcx* actually exhibits developmental and tissue-specific variation in escape from X-silencing (Carrel et al., 1996b, Sheardown et al., 1996). The degree of expression from the inactive X allele is highly variable and ranges from 20 % to 70 % of the active X allele levels. Imprinted extraembryonic tissues showed the most complete repression of the inactive X *Smcx* expression, while the highest levels of expression was seen in the heart (Carrel et al., 1996b).

Identification of male and female Coast moles

The success of the methylation-based X-inactivation assay, or any other method for that matter, depends on the reliable sexing of individual animals. The determination of sex in the Coast mole is complicated by the lack of any significant sexual dimorphism and the rather similar appearance of the external genitalia. The Coast mole is not unique regarding this aspect of their anatomy – the masculine appearance of the female European moles led researchers to believe that the animals were hermaphrodites (Jimenez et al., 1993, Sanchez et al., 1996a). However, there have been no indications that any North American mole possesses an ovotestis or exhibits masculinization to the degree seen in the European talpids. While it is possible to unequivocally determine the sex of the individual upon dissection, for many of our animals we were only provided with a DNA or tissue sample. Therefore, it was necessary to devise a DNA-based sexing method, such as one relying on the presence of the *Sry* gene, which is male-specific in all eutherians tested, including European moles (Sanchez et al., 1996b). The sequence of the *Sry* HMG box had been previously determined for the Algerian hedgehog (*Erinaceus algirus*), two shrew species (*Crocidura suaveolens*, *Neomys anomalus*) and three mole representative (*Talpa romana*, *T. europea*, *T. occidentalis*) (Sanchez et al., 1996b). The TeSRY1,2 primers was designed from the insectivore *Sry* HMG box sequence determined by Sanchez and colleagues (1996b) and were tested as indicators of sex in the Coast mole samples (Table 1).

Results

Sex determination using conserved Sry primers

PCR with *Sry* primers (TeSRY1,2) designed from the sequence of Sanchez et al. (1996b) was performed on all the available moles samples, some of which had already been sexed by K. Cambpell during dissections. Although slightly relaxed PCR conditions (50° C annealing, 40 cycles) were necessary to obtain strong bands, the primers worked extremely well and the PCR assay confirmed the sex of most of the animals, although in two cases the identification based on external morphology did not agree with the molecular results and was presumably incorrect (Figure 6).

Methylation analysis

Methylation-sensitive *HpaII* restriction digests followed by PCR with conserved primers spanning the CpG sites were performed for all three genes. *HindIII* digests were performed as negative controls and all the reactions were duplicated on murine DNA, which served as a positive control. Since the primers were heterologous with respect to the Coast mole template and the degree of sequence conservation was unknown, low stringency PCR conditions (50° C annealing, 40 cycles) were used. In the SMCX PCR, however, it was possible to use an annealing temperature of 60° C. Additionally, due to the CG richness of the template, increased concentrations of $MgCl_2$ and the addition of betaine were required (see Materials and Methods).

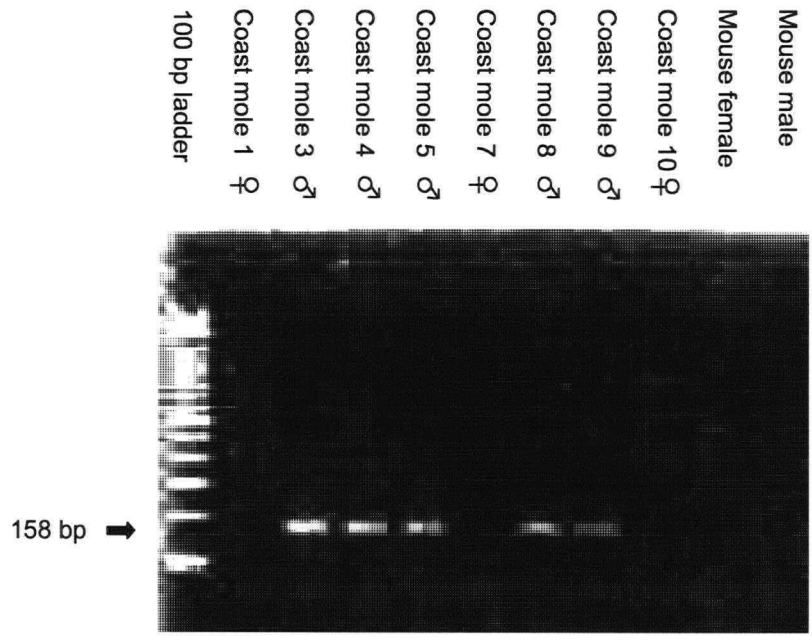


Figure 6: Coast mole DNAs amplified with the TeSRY1,2 primer pair
The 40 cycle PCR reaction was performed at an annealing temperature of 50C. The primers do not amplify mouse DNA, which served as a negative control in this reaction.

The *ZFX* primers, which were designed from conserved sequence at the human and murine CpG island (Luoh et al., 1995), amplify a 105 base pair fragment spanning two CCGG (*HpaII*) sites in mouse and humans (Table 3). Human CpG-island sequence was used to derive the *ARA* primers (Cutler Allen et al., 1992), which span two *HpaII* sites in both human and mouse. The *FMR1* primers, surrounding one *HpaII* cut site in humans and three in mouse, were similarly derived from human sequence (Fu et al., 1991). Both the *ARA* and *FMR1* CpG islands are associated with lengthy trinucleotide repeats (CAG in *ARA* and CGG in *FMR1*) that, when expanded, lead to the development of neurodegenerative disorders in humans (Strachan and Read, 1999).

ARA and *FMR1* fragments are only seen in the female Coast mole and female mouse control (see Figure 7), indicating that the CpG sites spanned by the primers were methylated in genomic DNA. The lack of a male *ARA* or *FMR1* signal is due to an absence of methylation at the same sites, which allowed the methylation-sensitive enzyme *HpaII* to cut the genomic template. The presence of bands in both male and female moles and mice in the *HindIII* lane confirms that the pattern seen in the samples digested by *HpaII* was not to lack of template or nonspecific digestion, but to a real methylation difference between the male and female animals at the *HpaII* sites. Thus, the female-specific *FMR1* and *ARA* bands seen in the Coast mole indicate that these genes are methylated and inactivated only in the females and not in the male mole and mouse. *ZFX*, on the other hand, shows a different pattern (Figure 7). A band is seen only in the female mouse samples, while the male mouse and male and female mole lanes are blank. Therefore, *Zfx* is methylated and inactivated only in the female mouse. The gene is unmethylated and, therefore presumed to be transcribed from the male mouse and from both sexes in the mole.

Attempts were made to perform this analysis on two other X-linked genes, *ALD1* and *SMCX*, but were unsuccessful. Although the *ALD1* primers (from Jegalian and Page, 1998) were chosen from CpG–island sequence conserved between human and mouse and appeared to work well in various eutherians (Jegalian and Page, 1998), it proved impossible to amplify a Coast mole fragment (data not shown).

The reasons for the failure of *SMCX* assay, on the other hand, were not due to the inability to amplify the Coast mole product. The *SMCX* primers, again designed from a conserved CpG island in mouse and humans (Jegalian and Page, 1998), not only amplified a female band following digestion with *HpaII*, but amplified a male band as well (Figure 7). Increasing the stringency of the PCR did not manage to eliminate this band. The inability of the methylation-sensitive restriction enzyme, *HpaII*, to cut the mole sequence was not due to the lack of *HpaII* sites since successful digests of the PCR products (which do not carry any methylation mark) demonstrated that these sites were present (data not shown). An additional problem with *SMCX* PCR was the persistence of a band in female mouse – a result that was at odds with published data (Jegalian and Page, 1998). The results of the methylation assay are summarized in Table 4.

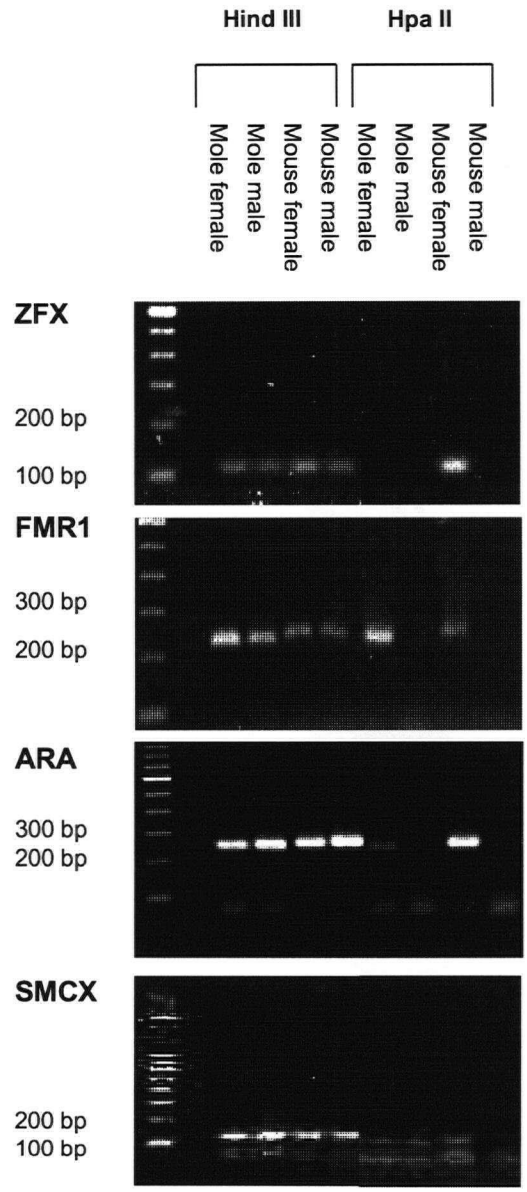


Figure 7. Methylation analysis of the *ZFX*, *ARA*, *FMR1* and *SMCX* genes in the Coast mole
Coast mole genomic DNA was digested with the methylation-sensitive enzyme *HpaII*, followed by PCR primers flanking the potentially methylated CpG sites. In the presence of methylation *HpaII* cannot cut and a PCR product is amplified. Since the methylation status of these genes in the mouse was known, mouse DNA and digestion with *HindIII*, which does not cut within the region amplified, served as controls.

	ARA		FMR1		SMCX		ZFX		Reference
	Amplification in methylation assay		Amplification in methylation assay		Amplification in methylation assay		Amplification in methylation assay		
Coast mole	F	M	F	M	F	M	F	M	This study.
	+	-	+	-	+	+	-	-	
	Inactivated		Inactivated		Inconclusive		Escapes		
Human	F	M	F	M	F	M	F	M	Jegalian and Page, 1998; Carrel and Willard, 1996a, Cutler-Allen et al., 1992.
	+	-	+	-	-	-	-	-	
	Inactivated		Inactivated		Escapes		Escapes		
Mouse	F	M	F	M	F	M	F	M	Confirmed in this study (except*); Jegalian and Page, 1998, Carrel et al., 1996b; Sheardown et al., 1996.
	+	-	+	-	+*	-	+	-	
	Inactivated		Inactivated		Inactivated *		Inactivated		

- Data are in contrast to published results (Jegalian and Page, 1998).

F - female, M - male

Table 4. Summary of methylation analysis.

Discussion

Sry has been identified as a male-specific gene, in a wide range of eutherians, including European moles and has been proven to be the male switch in the sex determination cascade in humans and mice (Sanchez et al., 1996b, Pask et al., 2000). *Sry* is also found on the Y chromosome of several marsupial species and sequence conservation at the HMG box has been noted (Foster et al., 1992). However, there is no functional evidence that *Sry* is involved in sex determination in the Metatheria as it seems to be ubiquitously expressed (Foster et al., 1992, Pask et al., 2000). The absence of an *Sry* homologue in the monotremes and the loss of *Sry* from several eutherian species (mole-voles (*Ellobius*), spiny rats) imply that other genes may control testis determination in some mammals. The recent identification of X and Y-linked marsupial homologues the of the human sex-reversing gene *ATRX* (alpha thalassemia, mental retardation on the X) is especially interesting (Pask et al., 2000).

In humans and mice the ancestral Y homologue of *ATRX* is no longer present in the Y chromosome, while in marsupials the *ATRY* is not only present, it is also expressed. The marsupial *ATRX* is expressed everywhere except in the developing male gonad, while the *ATRY* product is testis-specific. The *ATRX/ATRY* gene pair does not conform to the current model of evolution for the sex chromosomes (Lahn and Page, 1997). According to this theory, the Y carries two classes of genes with different evolutionary origins. The Class I genes arose from shared X/Y sequences and are expressed ubiquitously, while Class II is comprised of male-specific genes transposed onto the Y chromosome and exclusively expressed in the testis. Recent evidence has shown that even the genes with testis-specific expression may actually trace their origin to ancient XY gene pairs (Pask et al., 2000). Based on the fact that human

ATRX is involved in the sex-determination cascade downstream of *Sry* and several other genes and its expression pattern in marsupials, Pask and colleagues (2000) propose that *ATRX/ATRY* may represent the ancestral testis-determining mechanism. However, *Sry* is present only in the Coast mole males, indicating that it is borne on the Y chromosome in these organisms. Although in the absence of functional data it is impossible to say whether *Sry* is the testis-determining gene in moles, its presence on the Y chromosome is a very strong indicator that *Scapanus* conforms to the usual eutherian method of sex determination.

Despite the demanding PCR conditions necessary in the methylation analysis, it is still surprising that murine or human primers worked at all on an evolutionarily distant insectivore. However, the high level of conservation, the authors used the term "unprecedented", seen at the human and mouse *ZFX* CpG island indicates that there may be a selective pressure to conserve the 5' regulatory sequences of X-linked genes to ensure proper and stable inactivation (Luoh et al., 1995). The CpG-islands in the *ARA* and *FMR1* genes also seem to be relatively well conserved in the Eutheria, although the size of the associated trinucleotide repeats differs between species and between individuals (Cutler Allen et al., 1992, Eichler et al., 1995, Krongrad et al., 1995). In light of these results, it is unclear why the *ALD* PCR primers, designed from sequence conserved between humans and mice, was unsuccessful in moles, even though the gene is located on the conserved region of the X chromosome. It is possible that the region from which the primers had been selected was not as conserved as was initially thought and had diverged sufficiently in the mole to impede amplification.

The problems with the *SMCX* assay, in which a male and female PCR fragment were seen after digestion, did not allow us to make any firm conclusions regarding the inactivation status of the gene in the Coast mole. The existence of a *SMCY*, a housekeeping gene, has been

established in many eutherians, although it is not clear if the gene is functional in all cases, but its presence has not been studied in any insectivore (Jegalian and Page, 1998). Jegalian and Page (1998) reported CpG island methylation at the *SMCX* gene in the hedgehog, but did not comment on the presence of a Y-linked homologue. It is possible that moles have several *SMCY* copies, one of which may be functional, and that one of the gene copies is methylated, giving rise to the observed PCR fragment. In fact, a large proportion of Y chromosome is heterochromatic and consists of repeated sequences and pseudogenes, consistent with its degeneration after recombination with the X chromosome was suppressed (Chapter 1). However, multiple copies of *SMCY* have not been seen in other eutherians (Lahn et al., 2001) and one would not expect the inactivation of the X-linked gene if the Y homologue is functional. Alternatively, mole *SMCY* may have become nonfunctional and methylated, which in turn led of the silencing of the X-linked copy. In these two scenarios, the presence of a female band, despite the mysterious origin of the male fragment would still indicate that the gene is subject to X-inactivation. However, a third possibility is that the PCR bands were due to duplicated and silenced *SMCX/Y* pseudogenes that are autosomally located, in which case the results of the methylation assay are invalid and do not provide any information with respect to the inactivation of this gene. Nevertheless, the data allow us to conclude that at least one methylated *SMCX/Y*-like sequence exists in both male and female moles.

The presence of a *Smcx* PCR fragment in the female mouse does not agree with the data of Jegalian and Page (1998). This inconsistency may be due to several reasons, such as the tissue or strain source of the DNA or PCR conditions. Jegalian and Page (1998) used DNA extracted from the liver and spleen of a BALB/c mouse, while our sample originated from the spleen of a C57Black6 mouse. Since *Smcx* shows only partial escape from X-chromosome

inactivation, which differs between tissues and seems to change through development (Carrel et al., 1996, Sheardown et al., 1996), the DNA source tissue could have made a significant difference. On the other hand, since the mouse tissues used do not appear to differ strikingly, it is also possible that the degree and location of escape varies between the mouse strains. Additionally, the PCR conditions were not identical in the two experiments. For example, Jegalian and Page (1998) amplified only 50 ng of DNA using an annealing temperature of 62° C and 30 reaction cycles, while I used 100 ng of DNA, 60° C for annealing and 30 to 40 cycles of amplification. It is possible that the PCR under the conditions employed in this study was capable of amplifying very low levels of methylated *Smcx* sequences that were not detected by Jegalian and Page (1998).

The successful methylation analysis of *ARA*, *FMR1* and *ZFX* showed that *ARA* and *FMR1* are subject to silencing in the Coast mole, while *ZFX* escapes inactivation. In the mouse samples all three genes are inactivated, an observation that is in accordance with what is known about murine X-inactivation (Heard et al., 1997, Jegalian and Page, 1998). The fact that the inactivation states of the mouse genes, which served as a positive control, match those reported in the literature, confirms that the methylation assay is reliable method for investigating the question of silencing. The presence of sex-specific methylation in the Coast mole is a very strong indicator that the observed genes are sex-linked and that these animals undergo X-chromosome inactivation. The results of this assay, in combination with the methylation observed in the hedgehog (Jegalian and Page, 1998) show that silencing through the methylation of CpG islands is present in both of the lineages comprising the core Insectivores, the Soricomorphs and the Erinaceomorphs. Thus, methylation seems to be an integral part of X-inactivation in the Eutheria and must have been recruited into the dosage compensation

mechanism after the split with the Metatheria, in which methylation of CpG islands on the inactive X chromosome has not been observed (Kaslow and Migeon, 1987, Loebel and Johnstone, 1996). Additionally, the methylation analysis of *ZFX* presented in this thesis in combination with the data indicating *ZFX* is not silenced in hedgehogs (Jegalian and Page, 1998), confirms that this gene escapes inactivation in insectivores of the Erinaceomorph and Soricomorph lineage. In conclusion, the results of the methylation-sensitive restriction digest and PCR assay not only reveal that X-chromosome inactivation exists in *Scapanus orarius*, but also confirm that silencing through methylation is conserved in the insectivores and the Eutheria.

Chapter 4

Xist in the Coast mole: identification, expression and sequence analysis

Introduction

The *XIST/Xist* gene is the only gene expressed exclusively from the inactive X chromosome and encodes a large polyadenylated functional RNA (Brown et al., 1991, Brockdorff et al., 1992, Brown et al., 1992,). While the mechanism of *XIST/Xist* action remains unclear, there is no doubt that this RNA is absolutely essential for X-chromosome inactivation in eutherians. X chromosomes lacking the X inactivation centre (*XIC*), and thus the entire *XIST/Xist* gene cannot inactivate and neither can chromosomes with partial deletions of the gene (Marahrens et al., 1996, Penny et al., 1997). Gain-of-function experiments involving transgene constructs have confirmed that *Xist* is necessary and sufficient for inactivation, although the counting and choice aspect of X-chromosome inactivation seem to lie outside of the *Xist* gene (Avner and Heard, 2001) (for more detail refer to Chapter 1).

An *XIST* homologue has been identified in every eutherian species examined (cow, horse, rabbit, cat, rat, pig, various primates – see Appendix 1 for GenBank accession numbers), although expression studies have only been done in humans, mice, voles and cows. At this time there are no indications that a marsupial *Xist* genes has been identified (Wakefield, 1998). Until very recently the only complete sequences of the gene were those of human and mouse (Brown et al., 1991, Brockdorff et al., 1992, Brown et al., 1992), however on March 23rd, 2001 Nesterova et al. submitted the complete *Xist* sequences for three species of voles (Microtinae,

Rodentia): *Microtinus rossiameridionalis*, *M. argavis* and *M. transcaspicus*. Partial *Xist* sequences are also available for the horse, cow, rabbit, cat, pig, rat, rhesus monkey, African green monkey, howler monkey and chimpanzee. All of the partial sequences, except the rat, pig, cat and two bovine ESTs, encompass the 5' repeat and promoter region and were obtained by screening of cDNA libraries with a human *XIST* cDNA probe (Hendrich et al., 1993). The remaining cat, rat and bovine sequences are very short fragments obtained as ESTs or in genomic surveys. The available *Xist* sequences are shown in Figure 8.

Sequence conservation of the *XIST* gene amongst the examined species is generally low. The overall sequence similarity between the human and murine genes is only 60-70 %, although some regions show up to 80 % identity (Brown et al., 1992, Hendrich et al., 1993). Although several open reading frames (ORFs) were identified in the murine and human sequences, they were all quite small (483 bp in humans, 576 bp in mouse) and unlikely to be encoded by such a large transcript as *XIST/Xist* (Brockdorff et al., 1992, Brown et al., 1992). None of the ORFs were shared between the two species nor spanned an entire exon. Additionally, the best candidate for a protein-coding region in the human gene (ORF#4), was not well conserved in other primates (Brown et al., 1992, Hendrich et al., 1993). The combination of the lack of a conserved ORF and nuclear localization, has led to the widely accepted view that *XIST/Xist* is not translated and functions at the RNA level.

An unusual feature of *XIST* is the presence of several regions of direct tandem repeats. Two regions of these repeats are relatively conserved between mouse and human (the 5' repeats and downstream C-rich repeats), while others have been differentially amplified in the two genes. By performing Southern analysis with a variety of human *XIST* probes, Hendrich et al. (1993) identified the region of conserved repeats in the 5' region of exon one and located another 380

bp region of conservation approximately one kilobase downstream of the repeats within the same exon in several eutherian species. In the 380 bp area the human, lepine and bovine sequences, showed approximately 70 % identity. Mouse, on the other hand, exhibited only 47 % identity to the human sequence. In general, the alignment shows small regions of high similarity interspersed by areas of high divergence, as could be expected for a functional RNA in which secondary structure rather than primary sequence similarity is selected for. The primate *Xist* fragments were very similar, in some cases identical, to the human sequence. Similarly, the vole *Xist* genes, which also contain the 5' repeats described by Hendrich et al. (1993), and the small rat fragment seem to be very similar to the mouse, although the detailed vole sequence analysis is not yet available to the public (Nesterova et al., 2001 *in press*). Thus, despite its low sequence conservation, *Xist* is present in all eutherians studied, supporting its essential role in X-chromosome inactivation in this mammalian subclass.

The Coast mole is a eutherian mammal and, based on the results of the methylation assay presented in Chapter 3, this animal seems to utilize DNA methylation in X-linked gene silencing just like other eutherians. Although methylation of the inactive X chromosome is not seen in the Metatheria, methylated CpGs are present in other regions of the marsupial genome (Loebel et al., 1996, O'Neill et al., 1998). The lack of marsupial X-chromosome methylation has generally been thought of as an ancestral mammalian characteristic (Heard et al., 1997). However, since neither methylation nor X-chromosome inactivation have been extensively studied in the monotremes and are generally unknown (Cooper et al., 1993), the unlikely possibility remains that methylation of CpG islands on the X chromosome is the ancestral mammalian condition that was lost in the metatherians. In this scenario, the CpG island methylation of X-linked

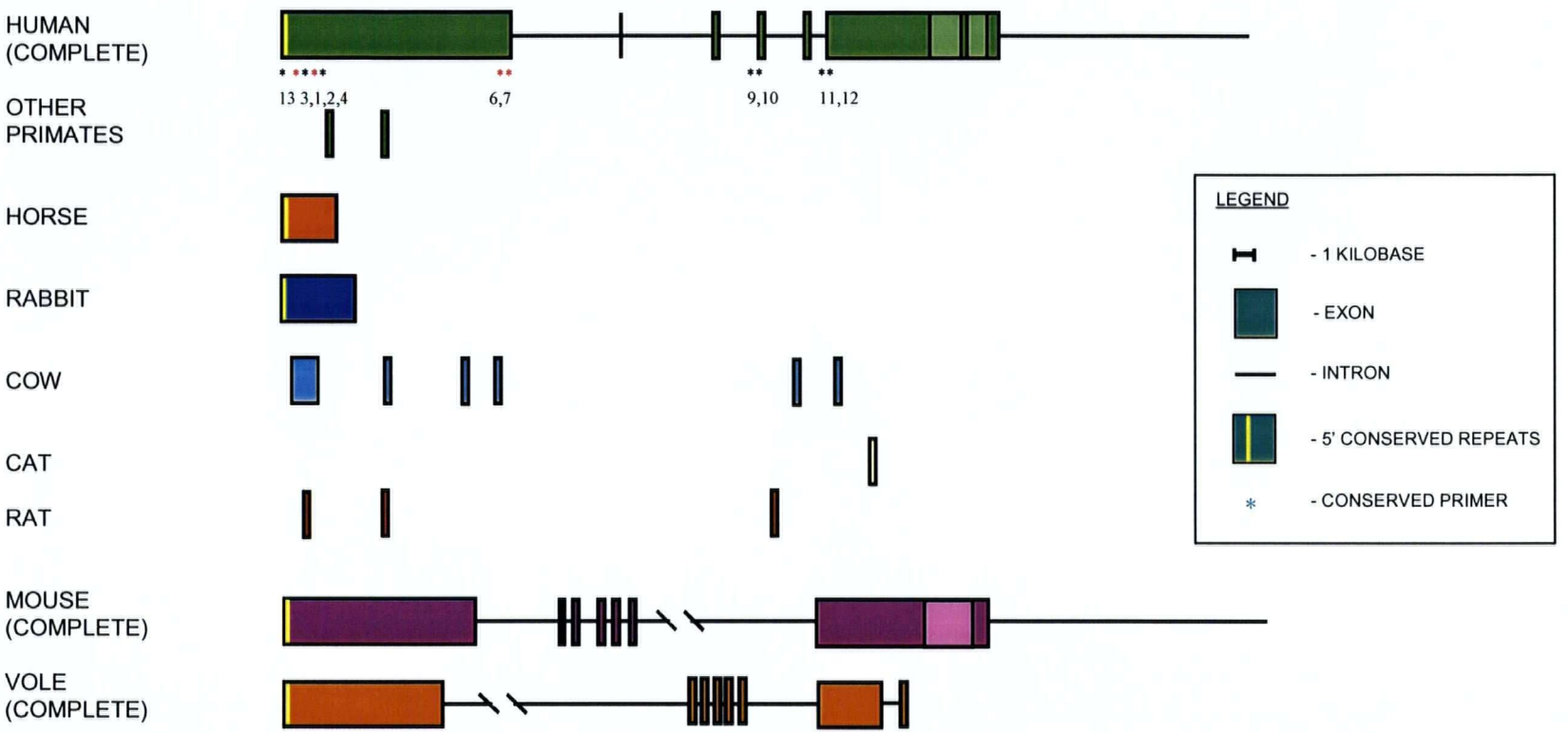


Figure 8: Available *Xist* sequences as of April 10th 2001 .
All of the fragments, except the rodents, are positioned based on their alignment with human *XIST*. The representative vole sequence (*Microtus rossiameridionalis*) and mouse gene have not been aligned with any other sequences. Lighter shades in the mouse and human genes indicate alternatively spliced products. Numerous ESTs fell within the same area and representative ones were used (See Appendix 1 for accession numbers). The approximate locations of conserved Coast mole primers are shown as asterisks (red for successful primers and black for unsuccessful primers).

sequences in *Scapanus* may represent an ancestral condition and may not mean that other aspects of the inactivation process in this species match those seen in mouse and humans. Apart from the methylation differences, another distinction between the inactive X in the two mammalian infraclasses is the presence of *Xist*, a gene that has been identified in all eutherians examined, but remains elusive in the marsupials (Hendrich et al., 1993, Wakefield, 1998). A crucial factor that would confirm that elements of X-chromosome inactivation are indeed conserved within the Eutheria would be the identification of a functional *Xist* homologue in the Coast mole.

Previous studies trying to detect *Xist* in species other than mouse and humans have used Southern analysis (Hendrich et al., 1993) or have localized the RNA by probing with heterologous probes, as was done by Duthie et al. (1999) in their study of vole chromosomes using a mouse *Xist* probe. PCR was not attempted or was unsuccessful, except when amplifying sequences from very closely related organisms, such as using human primers on primate sequences (Hendrich et al., 1993). However, in order to identify a gene by Southern analysis it is necessary to obtain a genomic or cDNA library of the organism of interest, neither of which exist for the mole nor any insectivore. A lack of high quality mole RNA early in the study prevented us from constructing our own cDNA library and led to the decision to attempt the identification of Coast mole *Xist* using PCR-based techniques – an approach that ultimately proved fruitful.

Results

Identification of Coast mole Xist

Before attempts were made to identify mole *Xist* through PCR it was necessary to obtain a positive PCR control. Since no suitable mole sequences were available in the database, primers had to be designed using sequence information from other species. *HPRT* (hypoxanthine-guanine phosphoribosyltransferae) cDNA sequences from a broad range of eutherian and one marsupial species, all available from GenBank, were aligned and regions of high conservation were identified. The primers SKHPRT1, 2 were picked manually and were expected to amplify both genomic and cDNA (see Table 3). However, although this primer pair works on human, bovine and mouse DNA and cDNA, it only seems to amplify mole cDNA. These results could indicate the presence of an intron between the primers in the mole. After this unexpected discovery, the SKHPRT1,2 primers were used to check for presence of mole cDNA. The rapid identification of a mole *Xist* PCR fragment from genomic DNA and the mole-specific primers designed from this fragment served as a positive DNA control.

Initially, I attempted, without any success, to detect mole *Xist* by using human and murine primers available in the laboratory in very low stringency PCR reactions (annealing temperature as low as 47° C, 40 cycles). Even the mXist1,2 primer pair, designed from exon 6 sequence conserved between mouse and human, failed to amplify a band in *Scapanus*. Keeping in mind the low identity score seen for mouse in the alignment of the 380 bp conserved fragment described by Hendrich et al. (1993), it seemed necessary to include more than the human and mouse sequences when designing primers expected to work across a range of species. Even if

mouse had not shown such poor alignment with the other sequences, it would still have been beneficial to expand the number of sequences aligned when designing the inter-specific primers to maximize the chances of identifying a region that was truly conserved.

The 5' region of exon one was aligned between human, bovine, lepine and equine sequences and regions of high similarity were identified. The mouse sequence was initially included in the alignment, but proved to be somewhat of an outlier showing significantly less conservation (between 11 % and 60 %) with respect to the other sequences in this region, which showed similarities of about 70 %. The removal of the murine sequence dramatically increased the quality of the alignment and facilitated primer choice. The locations and amplification conditions of the four primers selected, SKXIST1, SKXIST2, SKXIST3, SKXIST4, and other conserved primers are shown in Table 5 (for primer sequences see Figure 4 in Materials and Methods).

Although all four primers appeared to be well conserved between the species examined, only the SKXIST2 and SKXIST3 primer combination produced a 286 bp fragment in Coast mole DNA using relatively relaxed PCR conditions (50° C annealing temperature, 40 cycles), which was then gel-purified and sequenced. BLAST searches using this mole sequence against the non-redundant database detected good matches with all the known *Xist* sequences in this area of the gene, with the exception of mouse. CLUSTAL W (v. 1.8) alignments with the 245 base pair mole fragment (the 286 base pair fragment without the primer sequences), showed significant similarity to the human, cow, horse and rabbit *Xist* regions from which the primers were derived (Figure 9). The highest identity was with the horse sequence (81 %), the lowest with human at 59 %, while the cow and rabbit were 77 % and 76 % identical to the Coast mole

fragment. The comparatively low alignment score between the human gene and the remaining sequences is primarily due to stretch of C-rich repeats not present in the other species examined.

The mouse sequence showed very poor alignment in this region with 22 –59 % identity, while the voles, which were not included in the figure since they were not available at the time of this experiment, scored even lower. However, amongst themselves, the four rodent species exhibited a higher sequence identity. The three *Microtus* species were 98 % identity to each other and shared 70 % homology with the mouse.

Expression analysis of the putative Coast mole Xist fragment

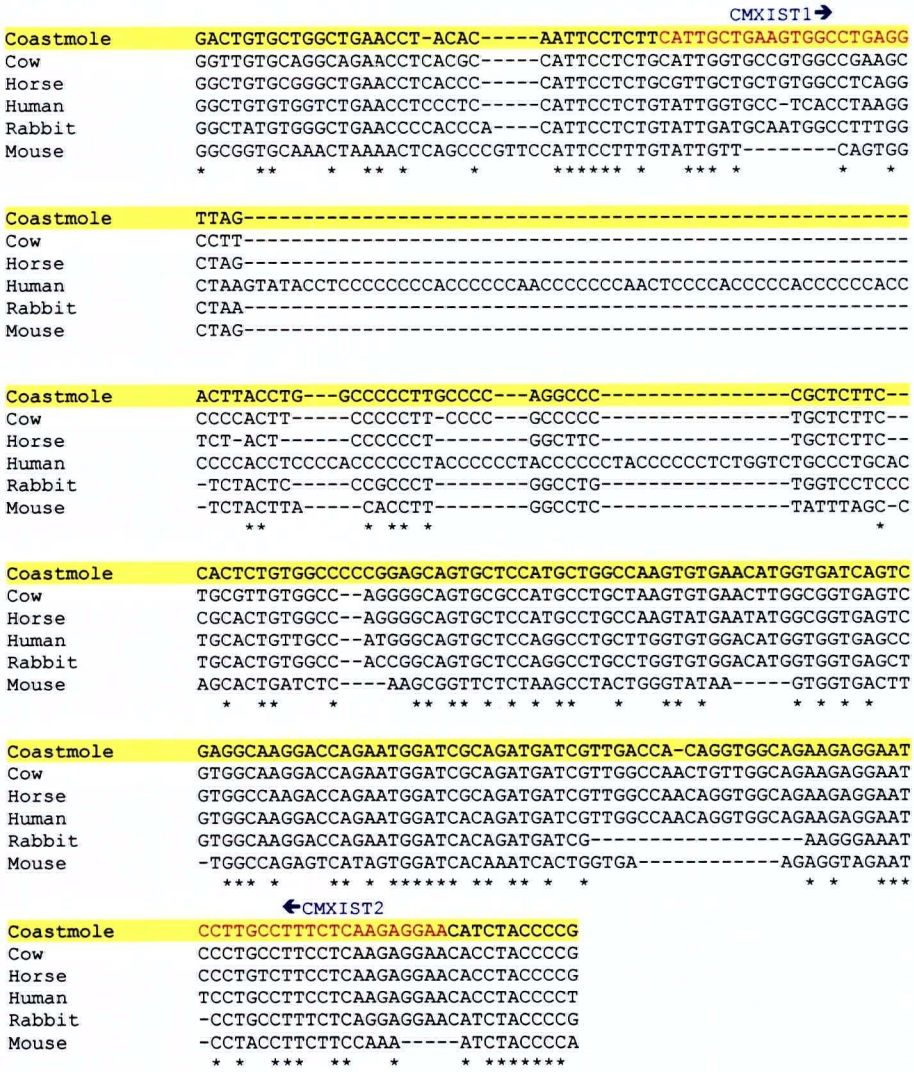
The sequence analysis of the newly identified 245 bp Coast mole band was a very good indicator that the fragment was indeed *Xist*. However, this sequence similarity, especially in a gene that is not especially conserved at the primary level, did not provide sufficient evidence for the existence of a Coast mole homologue. Additionally, even if the *Scapanus* fragment represented an *Xist*-related sequence, there was no evidence that it was functional and involved in X-chromosome inactivation in this species. If this Coast mole fragment was indeed *Xist*, one would expect it to follow the female-specific expression pattern seen for its homologues in mice, humans and cow.

Mole-specific primers CMXIST1 and 2 were designed from the 245 bp fragment, which allowed an increase in the stringency of the PCR conditions and more certainty when determining the expression status of the *Scapanus* sequence. However, before the expression of

Primer name	Location in human XIST (M97168)	Amplification in mole
SKXIST1	1446-1470	NO (with SKXIST2, SKXIST4)
SKXIST2	2255-2236	YES (with SKXIST3); 50° C, 40X
SKXIST3	1898-1908	YES (with SKXIST2); 50° C, 40X
SKXIST4	1913-1891	NO (with SKXIST1)
SKXIST6	29711-29688 (U80460 PAC, intronic)	YES (with SKXIST7); 50° C, 40X
SKXIST7	11010-11030	YES (with SKXIST6); 50° C, 40X
SKXIST9	11609-11628	NO (with SKXIST10)
SKXIST10	11915-11895	NO (with SKXIST9)
SKXIST11	13738-13760	NO (with SKXIST12)
SKXIST12	13892-13872	NO (with SKXIST11)
SKXIST13	3-23	NO (with CMXISTREV1, CMXIST2, CMXISTREV4, CMXIST7

Table 5. Conserved *Xist* primers.

The locations and sequences of the primers are shown in Figures 4, 11, 12 and 13.



PERCENT SEQUENCE SIMILARITY

	Coast mole	Cow	Horse	Human	Rabbit
Cow	79				
Horse	83	86			
Human	60	66	64		
Rabbit	78	77	82	63	
Mouse	29	22	57	43	59

Figure 9: The first Coast mole *Xist* fragment identified

The fragment is aligned with the human (M97168; bases 1913- 2234), bovine (AF104906; 1912-2153), lepine(U50910; 2782-3002), equine (U50911; 5679-5914) and murine genes (L04961; 2029-2237). The mole gene was amplified using the SKXIST2, 3 conserved primers, which lie outside of this alignment. The CMXIST1 and CMXIST2 mole-specific primers are indicated. The vole sequences were not available at the time of this experiment and are not included in the figure.

this putative *Xist* fragment could be examined, it was necessary to obtain undegraded RNA from the mole samples (Table 1). Although reasonably intact DNA could be obtained from frozen mole organs, animals found dead in traps or as "roadkill", the RNA from these specimens was of very poor quality. Undegraded RNA had to be obtained from recently sacrificed animals and from cell lines. Since no mole cell lines were available, kidneys from recently deceased Coast moles were used to establish the cell cultures. The cells, probably fibroblasts, initially grew well, but eventually started to senesce and differentiate, which is to be expected from a non-transformed cell line with a finite lifespan. The five female cultures (CM11, CM12, CM13, CM16, CM17) derived from sacrificed moles or animals killed in the field proliferated well. Unfortunately, the lone male cell line (CM15) did not grow as well, probably due to the fact that the mole from which the cell line was derived had been found in a trap and had been dead for an uncertain amount of time. RT-PCR reactions on the tissue sample and cell line RNAs were performed to create cDNAs, which could then be used as a template for PCR. Following PCR (54° C annealing, 40 cycles) with the CMXIST1 and CMXIST2 primers, the 245 base pair product was present in all the female cDNAs, but absent in the male CM15 samples (Figure 10).

Further sequencing of Coast mole Xist

The results of the previous two sections had established that *Xist* is present and probably functional in the Coast mole, since it is only expressed in the female. My next goal was to obtain as much *Scapanus Xist* sequence as possible to determine the size of the gene and to evaluate the global sequence conservation. Several more interspecific *Xist* primers were

designed from areas of the gene conserved between the available sequences - of these, only one primer pair was successful (Figure 4, Table 5). The identification of a short bovine EST at the very 3' end of human exon one and extending by 81 base pairs into the intron led to the design of the SKXIST6 and SKXIST7 primers, which managed to produce a 343 base pair band in the mole. This fragment was analyzed through BLAST and it showed significant identity to human and cow (74% and 75%). On the other hand, the SKXIST9 and 10 pair derived from a bovine EST corresponding to human exon four, amplified a band that, upon sequence analysis, proved to be either a PCR artifact or due to an unknown contaminant. BLAST searches with the SKXIST9,10 fragment failed to match any *Xist* sequences in the database, but did show a high similarity to a variety of cloning vectors. Additionally, it was impossible to align this fragment with the bovine EST and human sequence from which the SKXIST9,10 primers were derived (data not shown). Other primer pairs designed were SKXIST11 and 12, made from a short feline sequence fragment located in human exon six, while the alignment of murine, human, lepine and equine sequences served as a guide for the design of SKXIST13 at the start of exon one. None of these primers were successful in yielding a product after PCR.

Once the method of designing conserved primers was exhausted as a means for obtaining more Coast mole *Xist* sequence, I turned to inverse PCR. This technique, described in greater detail in Materials and Methods and Figure 3, allows for the amplification of sequence lying outside of the primers, which face away from each other, in contrast to conventional PCR. Amplification is achieved through restriction digestion and ligation reactions that result in a circular DNA molecule containing the two primers now facing each other and spanning the area of interest. Initially, two Coast mole-specific primer pairs facing away from each other were designed from the sequenced fragments (CMXIST1,2REVERSE and CMXIST5,6REVERSE).

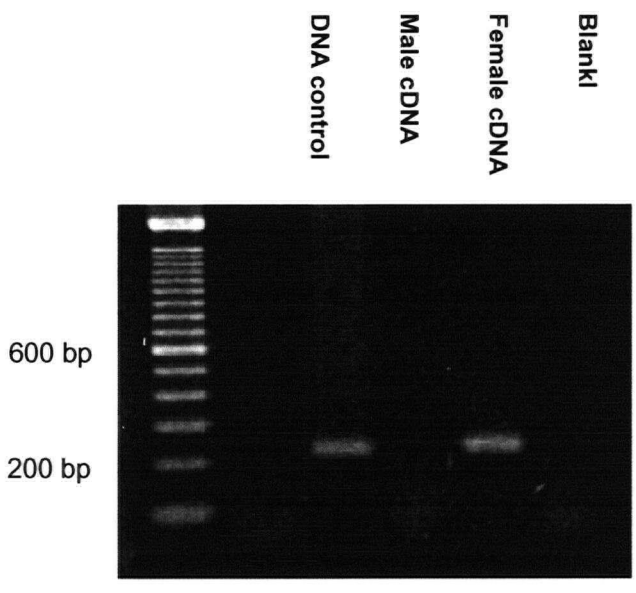


Figure 10: Female-specific expression of the putative Coast mole *Xist* fragment
PCR amplification with the CMXIST1 and CMXIST2 primer pair was successful only in cDNA derived from female animals or cell lines. PCR with SKHPRT1,2 primers was performed and confirmed the presence of RNA in both the male and female samples (data not shown).

Genomic DNA of Coast mole 10, was digested with enzymes that do not cut within or between the chosen primer sequences. The *AluI* and *RsaI* enzymes, both of which have a four base pair recognition site, were deemed as good choices for the digest since they should produce fragments of moderate size that would be amenable to PCR. Additionally, both enzymes left blunt ends upon digestion, which allowed double digests to be performed. Mole genomic DNA (Coast mole 10, a female) was then digested with each enzyme separately and in combination, ligated and ethanol precipitated. Initially the PCR reaction using *Taq* polymerase (Gibco) did not result in a band, however when the Expand Long PCR (Roche) enzyme kit was used, which is optimized for the creation of long PCR templates, large (1.2 kb and 1.0 kb, approximately) intensely-staining fragments were recovered for both the CMXIST1,2REVERSE and CMXIST5,6REVERSE primer pairs. These bands were gel purified, sequenced and contigs were assembled consisting of the original mole fragments through the use of conserved primers and the sequences obtained from inverse PCR. The Coast mole fragment at the 5' region of the gene had been increased to almost 1.4 kb, while the region at the end of exon one extended over 1.2 kb. Both of the sequenced regions of mole *Xist* have shown female-specific expression (see Figure 10 and data not shown). Additional inverse PCRs aimed at recovering the transcription initiation site at the beginning of exon one using the CMXIST6,7REVERSE primers and a variety of restriction enzymes have so far been unsuccessful. A schematic of all the Coast mole fragments obtained is presented in Figure 11.

Although the mole *Xist* sequence has been enlarged considerably, not much has changed with respect to sequence conservation. The mole fragment corresponding to the 5' region of exon one, shows between 68 % to 84 % homology with the rabbit, human, horse and bovine genes (Figure 12a,b), while the mouse and voles remain outliers with only about 20 % identity

to any of the sequences in this area. Different fragments in this region of the murine gene were used in order to optimize the alignment, but the scores remained very low. Similarly, the Coast mole fragment at the end of human exon one is 77 % similar to the human gene and shows poor alignment with the mouse and vole (approximately 40 %), while the identity between the murine and human sequence is 54 % (Figure 13a,b). Thus, the enlarged *Scapanus* fragments exhibit almost identical levels of homology as did the initial 286 and 343 base pair sequences. Note the while the alignments with and without the rodents in Figures 12 and 13 differ, the distance matrices (percent similarity) do not. This is a result of the CLUSTAL algorithm, which aligns the most closely related sequences first, which are then kept as a group, while more distantly related sequences (i.e. rodents) are added on without disrupting the relative alignments within the first group.

Even though the mouse and human *Xist/XIST* genes do not code for a protein product (Brockdorff et al., 1992, Brown et al., 1992, Hendrich et al., 1993), the available mole sequences were nevertheless checked for any ORFs using the GRAIL program (Uberbacher and Mural, 1991). The GRAIL algorithm requires one to choose a model organism in which to identify ORFs, and since mole was not available, the program was run twice, once using the human model and once with mouse model. The choice of organisms did not make a substantial difference in the detection of potential polypeptides. GRAIL identified seven very short ORFs, three of which were classified as having excellent coding potential, while three had good potential and one was of marginal quality. The size of the polypeptides encoded by these ORFs ranged between 7 and 35 amino acids. Five of the seven open reading frames overlapped with each other and encompassed the same area as ORF4 in the human sequence (Brown et al., 1992, Hendrich et al., 1993). No significant ORFs were identified in the 3' region of exon one.

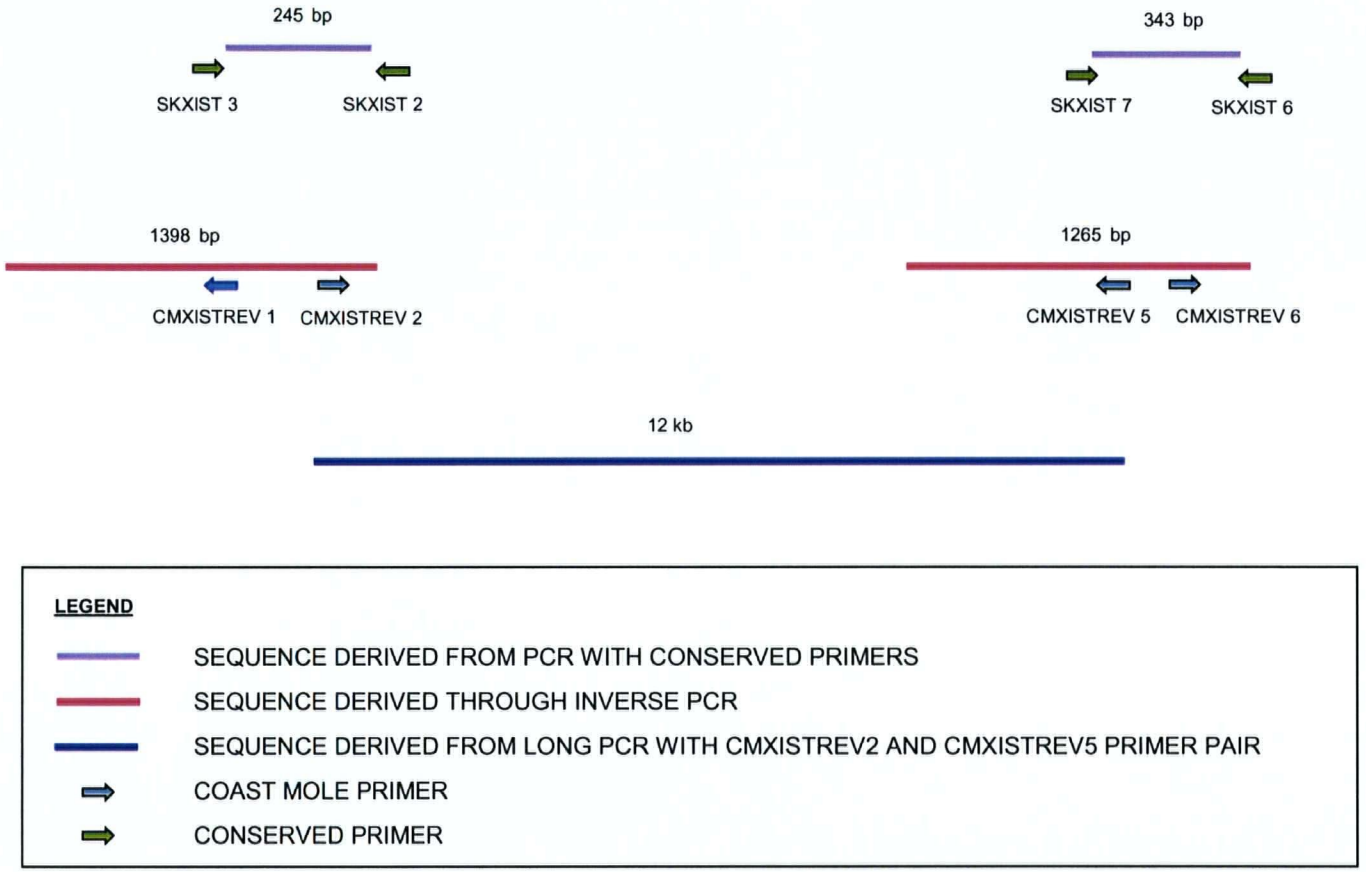
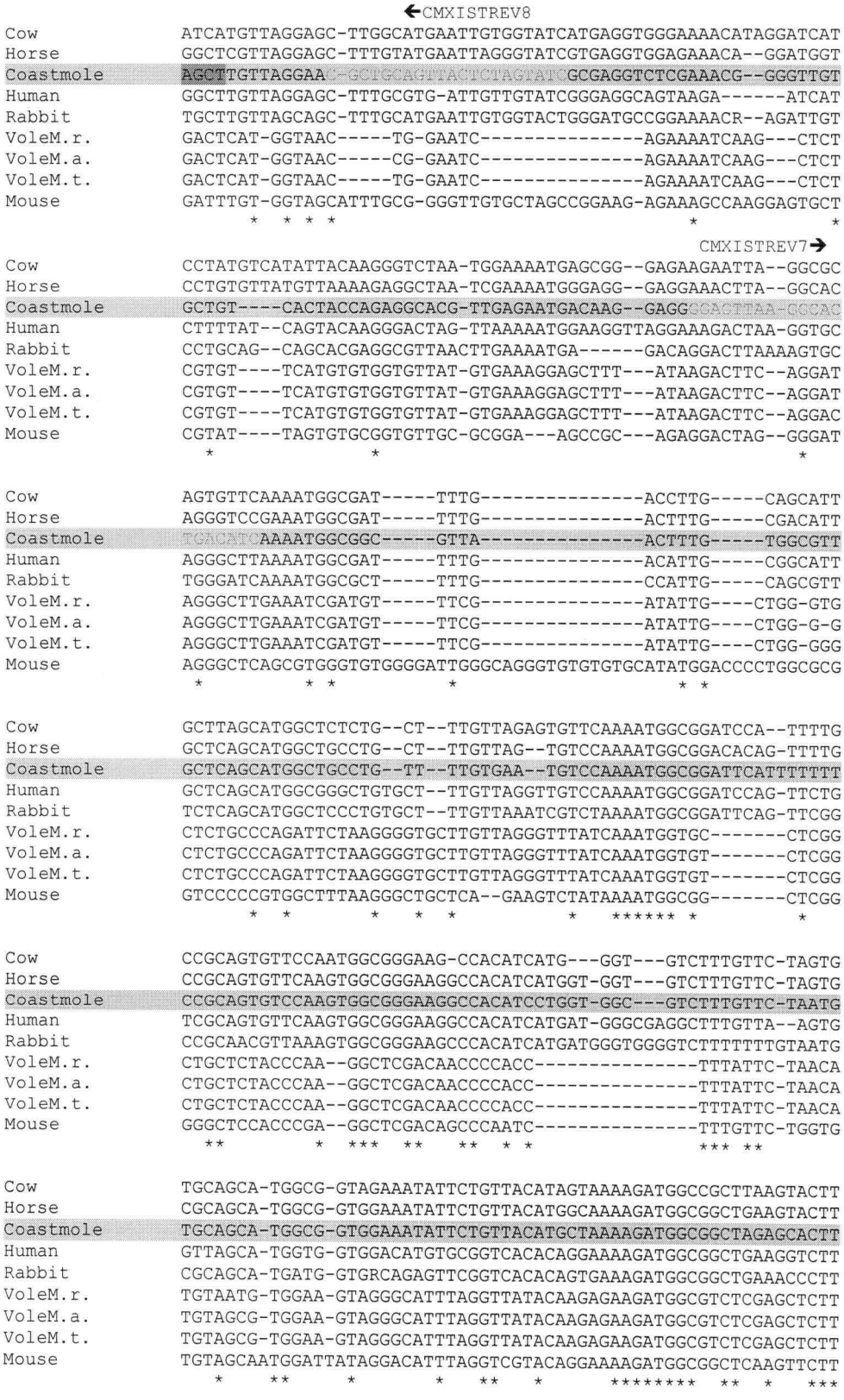
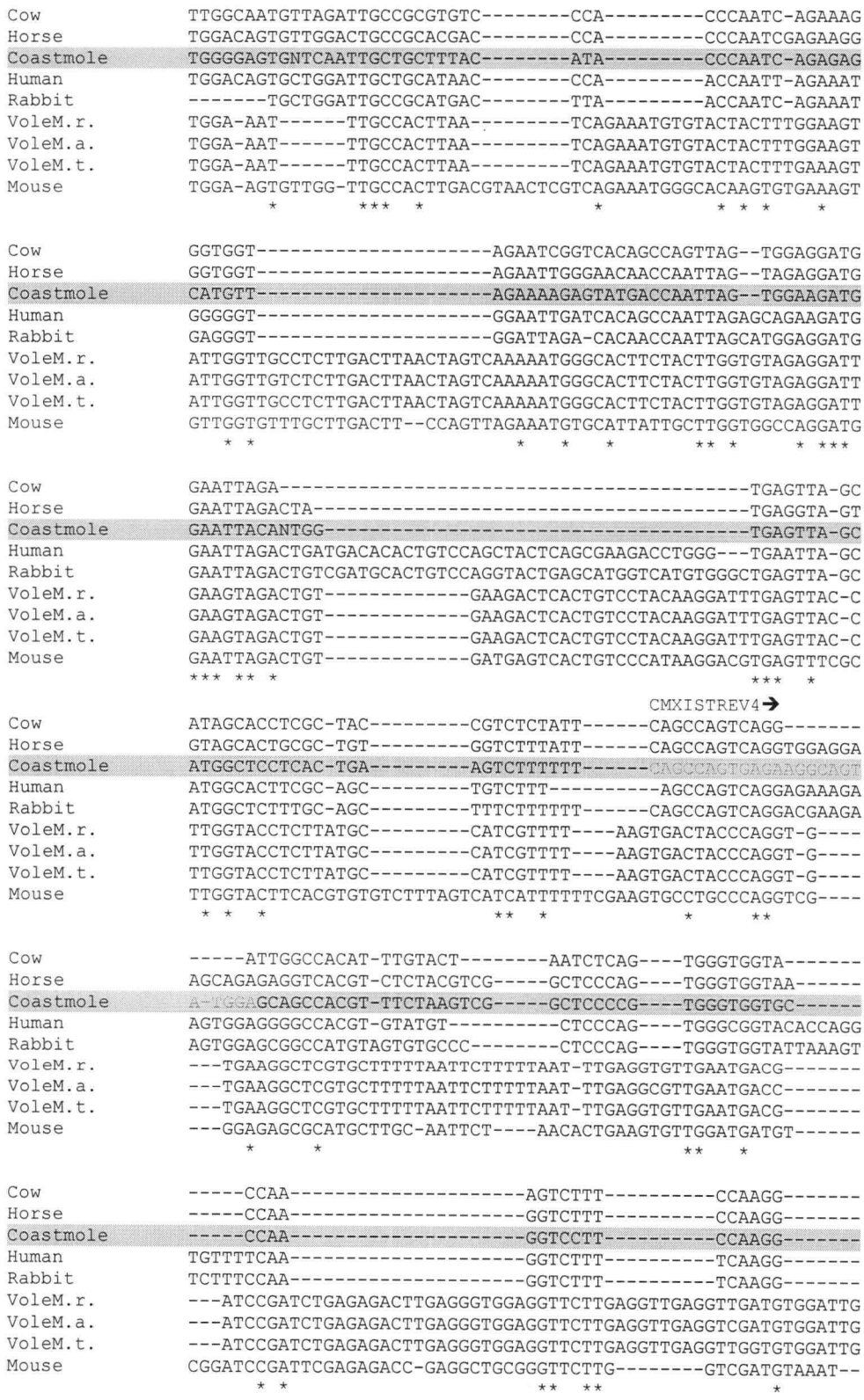
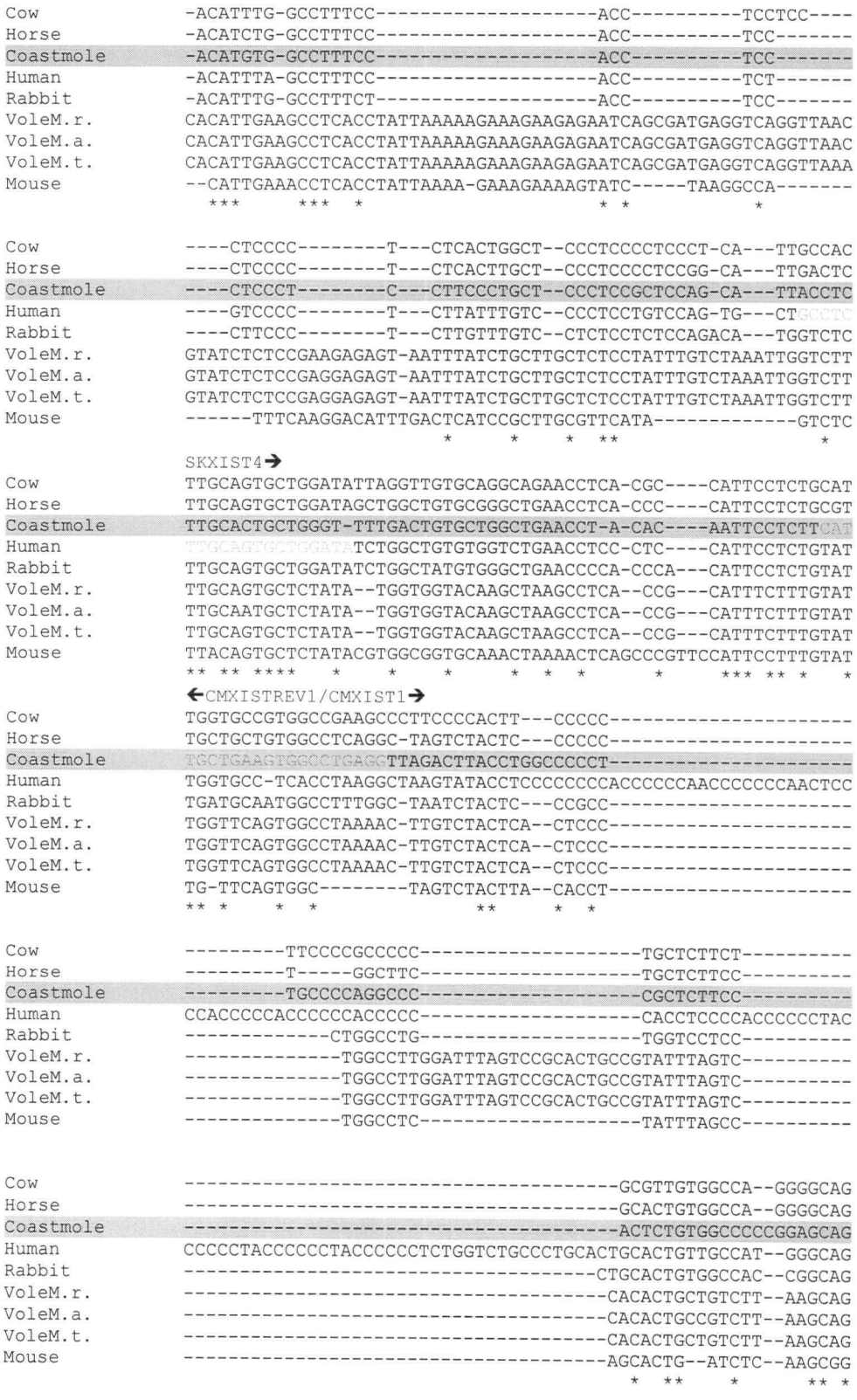


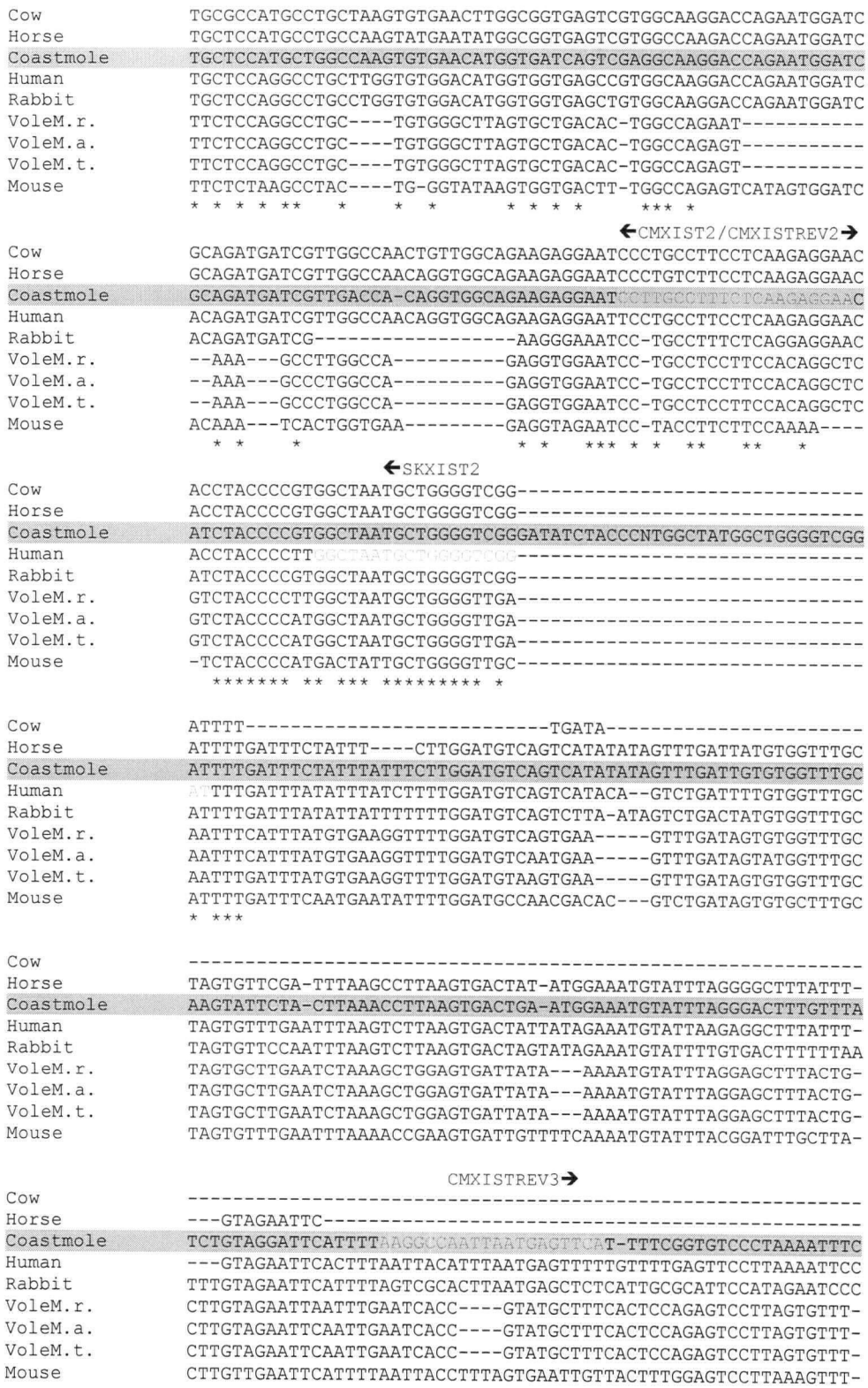
Figure 11: Coast mole *Xist* fragments.
The various Coast mole *Xist* sequences and the methods used to obtain them are summarized in this figure. The fragments are not drawn to scale and not all Coast mole and conserved primers are shown. For a complete list of primers see Tables 2,3 and 5 and Figure 4.


←CMXISTREV8
Cow ATCATGTTAGGAGC-TTGGCATGAATTGTGGTATCATGAGGTGGGAAAACATAGGATCAT
Horse GGCTCGTTAGGAGC-TTTGTATGAATTAGGGTATCGTGAGGTGGAGAAACA--GGATGGT
Coastmole AGCTTGTTAGGAA
Human GGCTTGTTAGGAGC-TTTGCGTG-ATTGTTGTATCGGGAGGCAGTAAGA------ATCAT
Rabbit TGCTTGTTAGCAGC-TTTGCATGAATTGTGGTACTGGGATGCCGGAAAACR--AGATTGT
VoleM.r. GACTCAT-GGTAAC-----TG-GAATC---------------AGAAAATCAAG---CTCT
VoleM.a. GACTCAT-GGTAAC-----CG-GAATC---------------AGAAAATCAAG---CTCT
VoleM.t. GACTCAT-GGTAAC-----TG-GAATC---------------AGAAAATCAAG---CTCT
Mouse GATTTGT-GGTAGCATTTGCG-GGGTTGTGCTAGCCGGAAG-AGAAAGCCAAGGAGTGCT
CMXISTREV7→
Cow CCTATGTCATATTACAAGGGTCTAA-TGGAAAATGAGCGG--GAGAAGAATTA--GGCGC
Horse CCTGTGTTATGTTAAAAGAGGCTAA-TCGAAAATGGGAGG--GAGGAAACTTA--GGCAC
Coastmole GCTGT----CACTACCAGAGGCACG-TTGAGAATGACAAG--GAGG
Human CTTTTAT--CAGTACAAGGGACTAG-TTAAAAATGGAAGGTTAGGAAAGACTAA-GGTGC
Rabbit CCTGCAG--CAGCACGAGGCGTTAACTTGAAAATGA------GACAGGACTTAAAAGTGC
VoleM.r. CGTGT----TCATGTGTGGTGTTAT-GTGAAAGGAGCTTT---ATAAGACTTC--AGGAT
VoleM.a. CGTGT----TCATGTGTGGTGTTAT-GTGAAAGGAGCTTT---ATAAGACTTC--AGGAT
VoleM.t. CGTGT----TCATGTGTGGTGTTAT-GTGAAAGGAGCTTT---ATAAGACTTC--AGGAC
Mouse CGTAT----TAGTGTGCGGTGTTGC-GCGGA---AGCCGC---AGAGGACTAG--GGGAT
Cow AGTGTTCAAAATGGCGAT-----TTTG---------------ACCTTG-----CAGCATT
Horse AGGGTCCGAAATGGCGAT-----TTTG---------------ACTTTG-----CGACATT
Coastmole AAAATGGCGGC-----GTTA---------------ACTTTG-----TGGCGTT
Human AGGGCTTAAAATGGCGAT-----TTTG---------------ACATTG-----CGGCATT
Rabbit TGGGATCAAAATGGCGCT-----TTTG---------------CCATTG-----CAGCGTT
VoleM.r. AGGGCTTGAAATCGATGT-----TTCG---------------ATATTG----CTGG-GTG
VoleM.a. AGGGCTTGAAATCGATGT-----TTCG---------------ATATTG----CTGG-G-G
VoleM.t. AGGGCTTGAAATCGATGT-----TTCG---------------ATATTG----CTGG-GGG
Mouse AGGGCTCAGCGTGGGTGTGGGGATTGGGCAGGGTGTGTGTGCATATGGACCCCTGGCGCG
Cow GCTTAGCATGGCTCTCTG--CT--TTGTTAGAGTGTTCAAAATGGCGGATCCA--TTTTG
Horse GCTCAGCATGGCTGCCTG--CT--TTGTTAG--TGTCCAAAATGGCGGACACAG-TTTTG
Coastmole GCTCAGCATGGCTGCCTG--TT--TTGTGAA--TGTCCAAAATGGCGGATTCATTTTTTT
Human GCTCAGCATGGCGGGCTGTGCT--TTGTTAGGTTGTCCAAAATGGCGGATCCAG-TTCTG
Rabbit TCTCAGCATGGCTCCCTGTGCT--TTGTTAAATCGTCTAAAATGGCGGATTCAG-TTCGG
VoleM.r. CTCTGCCCAGATTCTAAGGGGTGCTTGTTAGGGTTTATCAAATGGTGC-------CTCGG
VoleM.a. CTCTGCCCAGATTCTAAGGGGTGCTTGTTAGGGTTTATCAAATGGTGT-------CTCGG
VoleM.t. CTCTGCCCAGATTCTAAGGGGTGCTTGTTAGGGTTTATCAAATGGTGT-------CTCGG
Mouse GTCCCCCGTGGCTTTAAGGGCTGCTCA--GAAGTCTATAAAATGGCGG-------CTCGG
Cow CCGCAGTGTTCCAATGGCGGGAAG-CCACATCATG---GGT---GTCTTTGTTC-TAGTG
Horse CCGCAGTGTTCAAGTGGCGGGAAGGCCACATCATGGT-GGT---GTCTTTGTTC-TAGTG
Coastmole CCGCAGTGTCCAAGTGGCGGGAAGGCCACATCCTGGT-GGC---GTCTTTGTTC-TAATG
Human TCGCAGTGTTCAAGTGGCGGGAAGGCCACATCATGAT-GGGCGAGGCTTTGTTA--AGTG
Rabbit CCGCAACGTTAAAGTGGCGGGAAGCCCACATCATGATGGGTGGGGTCTTTTTTTGTAATG
VoleM.r. CTGCTCTACCCAA--GGCTCGACAACCCCACC---------------TTTATTC-TAACA
VoleM.a. CTGCTCTACCCAA--GGCTCGACAACCCCACC---------------TTTATTC-TAACA
VoleM.t. CTGCTCTACCCAA--GGCTCGACAACCCCACC---------------TTTATTC-TAACA
Mouse GGGCTCCACCCGA--GGCTCGACAGCCCAATC---------------TTTGTTC-TGGTG
Cow TGCAGCA-TGGCG-GTAGAAATATTCTGTTACATAGTAAAAGATGGCCGCTTAAGTACTT
Horse CGCAGCA-TGGCG-GTGGAAATATTCTGTTACATGGCAAAAGATGGCGGCTGAAGTACTT
Coastmole TGCAGCA-TGGCG-GTGGAAATATTCTGTTACATGCTAAAAGATGGCGGCTAGAGCACTT
Human GTTAGCA-TGGTG-GTGGACATGTGCGGTCACACAGGAAAAGATGGCGGCTGAAGGTCTT
Rabbit CGCAGCA-TGATG-GTGRCAGAGTTCGGTCACACAGTGAAAGATGGCGGCTGAAACCCTT
VoleM.r. TGTAATG-TGGAA-GTAGGGCATTTAGGTTATACAAGAGAAGATGGCGTCTCGAGCTCTT
VoleM.a. TGTAGCG-TGGAA-GTAGGGCATTTAGGTTATACAAGAGAAGATGGCGTCTCGAGCTCTT
VoleM.t. TGTAGCG-TGGAA-GTAGGGCATTTAGGTTATACAAGAGAAGATGGCGTCTCGAGCTCTT
Mouse TGTAGCAATGGATTATAGGACATTTAGGTCGTACAGGAAAAGATGGCGGCTCAAGTTCTT

```
Cow             GCCGCAGTCTAAAACATGGCGGGCTT--TTGTCTCTGCCGTGTGCATTTCCTGA-TAGGT
Horse           GCCGCAGTATAAAATATGGCGGGCTT--TTGTGTTTGCCGTGTGCATTTCCTGA-CAGGT
Coastmole       G--------TGTAACATGGCGG-CCT--TTGTCTTTGCTGTGTGTGTTTCCTGG-CAGGT
Human           GCCGCAGTGTAAAACATGGCGGGCCTCTTTGTCTTTGCTGTGTGCTTTTCGTGT-TGGGT
Rabbit          GCCGCAGTGTAAAATATGGCGGGCGTCTTTGTCTTTGCCGTGTGCATTTCGTGG-CGGGT
VoleM.r.        GCTGCAG--TATAAGGAAGCTGGCTT---TGTTTTTCACAGGCCAGTCTGGTGGCCAAAG
VoleM.a.        GCTGCAG--TATAAGGTAGCTGGCTT---TGTTTTTCACAGGCCAGTCTGGTGGCCAAAG
VoleM.t.        GCTGCAG--TATAAGGTAGCTGGCTT---TGTTTTTCACAGGCCAGTCTGGTGGCCAAAG
Mouse           GGTGCGG--TATAACGCAAAGGGCTT---TGTGTGTCACATGTCAGCTTCATG-TCTGAG
                *        *  **       * * *   *** * *     *      *  **

Cow             TTTGCTGCAGGGACAATATGGCTGACCTTGTCA--TGTGGATATCATGGCAGTTTGTCAC
Horse           TTTGCCGCAGGGACAATATGGCAGACCTTGTCA--TGTGGATATCATGGCAGTTCGTCAC
Coastmole       TTTGCCTCGGGGATAGTATGGCAGGCCTTATTA--TGTGGAAGTCTTGGCAATTTGTCAC
Human           TTTGCCGCAGGGACAATATGGCAGGCGTTGTCA--TATGTATATCATGGCT-TTTGTCAC
Rabbit          TTTGCCGCAGGGACAATATGGCTGACGATATCA--TGTGGTTATCATGGAGATTTGTCAC
VoleM.r.        TTAGCCTAGAGGACAATGTAGCACATGCTTTGGG--GTGGCCAAGATGGCG--ATGTCAC
VoleM.a.        TTAGCCTAGAGGACAATGTAGCACATGCTTTGGG--GTGGCCAAGATGGCG--ATGTCAC
VoleM.t.        TTAGCCTAGAGGACAATGTAGCACATGCTTTGGG--GTGGCCAAGATGGCG--ATGTCAC
Mouse           TTAGCCTGGAGAG----GTGGCACATGCTCTTGAATGTGTCTAAGATGGCGG-AAGTCAT
                ** **     *       * **      * *      **       ***      ****

Cow             GTGGAT-----------------------------------------------ATCGTGGCAG-
Horse           GTGGAC-----------------------------------------------GTCATGGCGG-
Coastmole       GTGGAT-----------------------------------------------GTCTTGGCAGA
Human           GTGGAC-----------------------------------------------ATCATGGC---
Rabbit          GTGGAC-----------------------------------------------GTCATGGCGC-
VoleM.r.        GTGATTGTCCCTTGAGGG-TTTTTTTCTGATATATTCTTGGCGGGCTTAGTCTTGGCGGG
VoleM.a.        GTGATTGTCCCTTGAGGG-TTTTTTTCTGATATATTCTTGGCGGGCTTAGTCTTGGCGGG
VoleM.t.        GTGATTGTCCCTTGAGGGCTTTTTTTCTGATATAGTCTTGGCGGGCTTAGTCTTGGCGGG
Mouse           GTGACCTGCCCTCTAGTGGTTTCTTTCAG------------------TGATTTT------
                ***                                               *  *

Cow             -------------------------GGGTGTTT-GACCGTTACATTCTTGGCGGGCTTTG
Horse           -------------------------GGGGATTTTGCCCGGTACATTCTTGGCGGGCTTTG
Coastmole       GGGTC---------------CACAGGGGCGTTC-GTCTG-CATGTTCTTGGCGGGCTCAG
Human           ------------------------------------------------------GGGCT---
Rabbit          ----------------------------GTTTTTGCTCGGTAAATT---GGCGGGCTTTG
VoleM.r.        CTTAGTCTTGGCGGGCTTAATCTTGGCGGGCTTTGCTTAGC-CACACATGGCGGGCTTTG
VoleM.a.        CTTA---------------ATCTTGGCGGGCTTTGCTTAGC-CACGCATGGCGGGCTTTG
VoleM.t.        CTTAATCTTGGCGGGCTTAATCTTGGCGGGCTTTGCTTAGC-CACACATGGCGGGCTTTG
Mouse           -TT-----------------TTTTGGCGGGCTT----TAGC-TAC--TTGGCGGGCTTTG
                                                                 *****

Cow             CATC------AGGAGGGCCTGCCGCATT--GTTAAAG-ATGGCG----------------
Horse           CACC------AGGTGGGCCTGCCGCATT--GTTAAAG-ATGGCGGGGTTTGGCGCGTTTT
Coastmole       CACC------TGGTGGTATTGCCGCATT--GTTAAAG-ATGGCGG---------------
Human           -------------------TGCCGCATT--GTTAAAG-ATGGCGG---------------
Rabbit          CACA------AGGTGGACCTACGGCATT--GTTAAAG-ATGGCGG---------------
VoleM.r.        TCCTTGTCCAAGATGTACTTAGTGCATTAAGATGTGA-GTGCTTG---------------
VoleM.a.        TCCTTGTCCAAGATGTACTTAGTGCATTAAGATGTGA-GTGCTTG---------------
VoleM.t.        TCCTTGTCCAAGATGTACTTAGTGCATTAAGATGTGA-GTGGTTG---------------
Mouse           CCC-----GAGGGTACACTTGGTGCATTATGGTAGGGTGTGGTTG---------------
                              *   *****  * *      **

Cow             ----TGCTTTGCCGCG-------------GAC----AAAGTG----AAAGGA----GGGA
Horse           GGCGCGTTTTGCCGCGTTTGCTGCGTTTTGCCCCGAAAAGTG----GAAGAAAGGTGGGG
Coastmole       -----GCTTTGCCGCGTTTG---------TGGC--GGAAAGCNCGAAGGAGAAACATGGGG
Human           -----GTTTTGCCGCCTAG---------TGCCACGCAGAGCG---GGAGAAAAGGTGGGA
Rabbit          -----GCGTTGCCGCATTT----------TGCCGGTCAAAGCA----GGAGAAAACCGGG-
VoleM.r.        -----GTTTTACCTTGT------------GCCACTCGAAGC--------------CCAGT
VoleM.a.        -----GTTTTACCTTGT------------GCCACTCGAAGC--------------CCAGT
VoleM.t.        -----GTTTTACCTTGT------------GCCACTCGAAGC--------------CCAGT
Mouse           -----GTCCTACCTTGT------------GCCACTCGAAGCT----GAGGCAAGGCTAAG
                     *   * **                * *      **
```


Cow TTGGCAATGTTAGATTGCCGCGTGTC--------CCA---------CCCAATC-AGAAAG
Horse TGGACAGTGTTGGACTGCCGCACGAC--------CCA---------CCCAATCGAGAAGG
Coastmole TGGGGAGTGNTCAATTGCTGCTTTAC--------ATA---------CCCAATC-AGAGAG
Human TGGACAGTGCTGGATTGCTGCATAAC--------CCA---------ACCAATT-AGAAAT
Rabbit -------TGCTGGATTGCCGCATGAC--------TTA---------ACCAATC-AGAAAT
VoleM.r. TGGA-AAT------TTGCCACTTAA---------TCAGAAATGTGTACTACTTTGGAAGT
VoleM.a. TGGA-AAT------TTGCCACTTAA---------TCAGAAATGTGTACTACTTTGGAAGT
VoleM.t. TGGA-AAT------TTGCCACTTAA---------TCAGAAATGTGTACTACTTTGAAAGT
Mouse TGGA-AGTGTTGG-TTGCCACTTGACGTAACTCGTCAGAAATGGGCACAAGTGTGAAAGT
GGTGGT----------AGAATCGGTCACAGCCAGTTAG--TGGAGGATG
GGTGGT----------AGAATTGGGAACAACCAATTAG--TAGAGGATG
CATGTT----------AGAAAAGAGTATGACCAATTAG--TGGAAGATG
GGGGGT----------GGAATTGATCACAGCCAATTAGAGCAGAAGATG
GAGGGT----------GGATTAGA-CACAACCAATTAGCATGGAGGATG
ATTGGTTGCCTCTTGACTTAACTAGTCAAAAATGGGCACTTCTACTTGGTGTAGAGGATT
ATTGGTTGTCTCTTGACTTAACTAGTCAAAAATGGGCACTTCTACTTGGTGTAGAGGATT
ATTGGTTGCCTCTTGACTTAACTAGTCAAAAATGGGCACTTCTACTTGGTGTAGAGGATT
GTTGGTGTTTGCTTGACTT--CCAGTTAGAAATGTGCATTATTGCTTGGTGGCCAGGATG
GAATTAGA----------TGAGTTA-GC
GAATTAGACTA----------TGAGGTA-GT
GAATTACANTGG----------TGAGTTA-GC
GAATTAGACTGATGACACACTGTCCAGCTACTCAGCGAAGACCTGGG---TGAATTA-GC
GAATTAGACTGTCGATGCACTGTCCAGGTACTGAGCATGGTCATGTGGGCTGAGTTA-GC
GAAGTAGACTGT----------GAAGACTCACTGTCCTACAAGGATTTGAGTTAC-C
GAAGTAGACTGT----------GAAGACTCACTGTCCTACAAGGATTTGAGTTAC-C
GAAGTAGACTGT----------GAAGACTCACTGTCCTACAAGGATTTGAGTTAC-C
GAATTAGACTGT----------GATGAGTCACTGTCCCATAAGGACGTGAGTTTCGC
CMXISTREV4➔
ATAGCACCTCGC-TAC---------CGTCTCTATT------CAGCCAGTCAGG-------
GTAGCACTGCGC-TGT---------GGTCTTTATT------CAGCCAGTCAGGTGGAGGA
ATGGCTCCTCAC-TGA---------AGTCTTTTTT------CAGCCAGTGAGAAGGCAGT
ATGGCACTTCGC-AGC---------TGTCTTT----------AGCCAGTCAGGAGAAAGA
ATGGCTCTTTGC-AGC---------TTTCTTTTTT------CAGCCAGTCAGGACGAAGA
TTGGTACCTCTTATGC---------CATCGTTTT----AAGTGACTACCCAGGT-G----
TTGGTACCTCTTATGC---------CATCGTTTT----AAGTGACTACCCAGGT-G----
TTGGTACCTCTTATGC---------CATCGTTTT----AAGTGACTACCCAGGT-G----
TTGGTACTTCACGTGTGTCTTTAGTCATCATTTTTTCGAAGTGCCTGCCCAGGTCG----
-----ATTGGCCACAT-TTGTACT--------AATCTCAG----TGGGTGGTA-------
AGCAGAGAGGTCACGT-CTCTACGTCG-----GCTCCCAG----TGGGTGGTAA------
A-TGGAGCAGCCACGT-TTCTAAGTCG-----GCTCCCCG----TGGGTGGTGC------
AGTGGAGGGGCCACGT-GTATGT----------CTCCCAG----TGGGCGGTACACCAGG
AGTGGAGCGGCCATGTAGTGTGCCC--------CTCCCAG----TGGGTGGTATTAAAGT
---TGAAGGCTCGTGCTTTTTAATTCTTTTTAAT-TTGAGGTGTTGAATGACG-------
---TGAAGGCTCGTGCTTTTTAATTCTTTTTAAT-TTGAGGCGTTGAATGACC-------
---TGAAGGCTCGTGCTTTTTAATTCTTTTTAAT-TTGAGGTGTTGAATGACG-------
---GGAGAGCGCATGCTTGC-AATTCT----AACACTGAAGTGTTGGATGATGT------
-----CCAA----------AGTCTTT----------CCAAGG-------
-----CCAA----------GGTCTTT----------CCAAGG-------
-----CCAA----------GGTCCTT----------CCAAGG-------
TGTTTTCAA----------GGTCTTT----------TCAAGG-------
TCTTTCCAA----------GGTCTTT----------TCAAGG-------
---ATCCGATCTGAGAGACTTGAGGGTGGAGGTTCTTGAGGTTGAGGTTGATGTGGATTG
---ATCCGATCTGAGAGACTTGAGGGTGGAGGTTCTTGAGGTTGAGGTCGATGTGGATTG
---ATCCGATCTGAGAGACTTGAGGGTGGAGGTTCTTGAGGTTGAGGTTGGTGTGGATTG
CGGATCCGATTCGAGAGACC-GAGGCTGCGGGTTCTTG--------GTCGATGTAAAT--


Cow -ACATTTG-GCCTTTCC--------ACC----------TCCTCC----
Horse -ACATCTG-GCCTTTCC--------ACC----------TCC-------
Coastmole -ACATGTG-GCCTTTCC--------ACC----------TCC-------
Human -ACATTTA-GCCTTTCC--------ACC----------TCT-------
Rabbit -ACATTTG-GCCTTTCT--------ACC----------TCC-------
VoleM.r. CACATTGAAGCCTCACCTATTAAAAAGAAAGAAGAGAATCAGCGATGAGGTCAGGTTAAC
VoleM.a. CACATTGAAGCCTCACCTATTAAAAAGAAAGAAGAGAATCAGCGATGAGGTCAGGTTAAC
VoleM.t. CACATTGAAGCCTCACCTATTAAAAAGAAAGAAGAGAATCAGCGATGAGGTCAGGTTAAA
Mouse --CATTGAAACCTCACCTATTAAAA-GAAAGAAAAGTATC-----TAAGGCCA-------
Cow ----CTCCCC--------T---CTCACTGGCT--CCCTCCCCTCCCT-CA---TTGCCAC
Horse ----CTCCCC--------T---CTCACTTGCT--CCCTCCCCTCCGG-CA---TTGACTC
Coastmole ----CTCCCT--------C---CTTCCCTGCT--CCCTCCGCTCCAG-CA---TTACCTC
Human ----GTCCCC--------T---CTTATTTGTC--CCCTCCTGTCCAG-TG---CTGCCTC
Rabbit ----CTTCCC--------T---CTTGTTTGTC--CTCTCCTCTCCAGACA---TGGTCTC
VoleM.r. GTATCTCTCCGAAGAGAGT-AATTTATCTGCTTGCTCTCCTATTTGTCTAAATTGGTCTT
VoleM.a. GTATCTCTCCGAGGAGAGT-AATTTATCTGCTTGCTCTCCTATTTGTCTAAATTGGTCTT
VoleM.t. GTATCTCTCCGAGGAGAGT-AATTTATCTGCTTGCTCTCCTATTTGTCTAAATTGGTCTT
Mouse ------TTTCAAGGACATTTGACTCATCCGCTTGCGTTCATA------------GTCTC
SKXIST4➔
Cow TTGCAGTGCTGGATATTAGGTTGTGCAGGCAGAACCTCA-CGC----CATTCCTCTGCAT
Horse TTGCAGTGCTGGATAGCTGGCTGTGCGGGCTGAACCTCA-CCC----CATTCCTCTGCGT
Coastmole TTGCACTGCTGGGT-TTTGACTGTGCTGGCTGAACCT-A-CAC----AATTCCTCTTCAT
Human TTGCAGTGCTGGATATCTGGCTGTGTGGTCTGAACCTCC-CTC----CATTCCTCTGTAT
Rabbit TTGCAGTGCTGGATATCTGGCTATGTGGGCTGAACCCCA-CCCA---CATTCCTCTGTAT
VoleM.r. TTGCAGTGCTCTATA--TGGTGGTACAAGCTAAGCCTCA--CCG---CATTTCTTTGTAT
VoleM.a. TTGCAATGCTCTATA--TGGTGGTACAAGCTAAGCCTCA--CCG---CATTTCTTTGTAT
VoleM.t. TTGCAGTGCTCTATA--TGGTGGTACAAGCTAAGCCTCA--CCG---CATTTCTTTGTAT
Mouse TTACAGTGCTCTATACGTGGCGGTGCAAACTAAAACTCAGCCCGTTCCATTCCTTTGTAT
←CMXISTREV1/CMXIST1➔
Cow TGGTGCCGTGGCCGAAGCCCTTCCCCACTT---CCCCC----------
Horse TGCTGCTGTGGCCTCAGGC-TAGTCTACTC---CCCCC----------
Coastmole TGCTGAAGTGGCCTGAGGTTAGACTTACCTGGCCCCCT----------
Human TGGTGCC-TCACCTAAGGCTAAGTATACCTCCCCCCCCACCCCCCAACCCCCCCAACTCC
Rabbit TGATGCAATGGCCTTTGGC-TAATCTACTC---CCGCC----------
VoleM.r. TGGTTCAGTGGCCTAAAAC-TTGTCTACTCA--CTCCC----------
VoleM.a. TGGTTCAGTGGCCTAAAAC-TTGTCTACTCA--CTCCC----------
VoleM.t. TGGTTCAGTGGCCTAAAAC-TTGTCTACTCA--CTCCC----------
Mouse TG-TTCAGTGGC--------TAGTCTACTTA--CACCT----------
Cow ---------TTCCCCGCCCCC----------TGCTCTTCT----------
Horse ---------T-----GGCTTC----------TGCTCTTCC----------
Coastmole ---------TGCCCCAGGCCC----------CGCTCTTCC----------
Human CCACCCCCACCCCCCACCCCC----------CACCTCCCCACCCCCCTAC
Rabbit ----------CTGGCCTG----------TGGTCCTCC----------
VoleM.r. ----------TGGCCTTGGATTTAGTCCGCACTGCCGTATTTAGTC----------
VoleM.a. ----------TGGCCTTGGATTTAGTCCGCACTGCCGTATTTAGTC----------
VoleM.t. ----------TGGCCTTGGATTTAGTCCGCACTGCCGTATTTAGTC----------
Mouse ----------TGGCCTC----------TATTTAGCC----------
Cow ----------GCGTTGTGGCCA--GGGGCAG
Horse ----------GCACTGTGGCCA--GGGGCAG
Coastmole ----------ACTCTGTGGCCCCCGGAGCAG
Human CCCCCTACCCCCCTACCCCCCTCTGGTCTGCCCTGCACTGCACTGTTGCCAT--GGGCAG
Rabbit ----------CTGCACTGTGGCCAC--CGGCAG
VoleM.r. ----------CACACTGCTGTCTT--AAGCAG
VoleM.a. ----------CACACTGCCGTCTT--AAGCAG
VoleM.t. ----------CACACTGCTGTCTT--AAGCAG
Mouse ----------AGCACTG--ATCTC--AAGCGG


Cow TGCGCCATGCCTGCTAAGTGTGAACTTGGCGGTGAGTCGTGGCAAGGACCAGAATGGATC
Horse TGCTCCATGCCTGCCAAGTATGAATATGGCGGTGAGTCGTGGCCAAGACCAGAATGGATC
Coastmole TGCTCCATGCTGGCCAAGTGTGAACATGGTGATCAGTCGAGGCAAGGACCAGAATGGATC
Human TGCTCCAGGCCTGCTTGGTGTGGACATGGTGGTGAGCCGTGGCAAGGACCAGAATGGATC
Rabbit TGCTCCAGGCCTGCCTGGTGTGGACATGGTGGTGAGCTGTGGCAAGGACCAGAATGGATC
VoleM.r. TTCTCCAGGCCTGC----TGTGGGCTTAGTGCTGACAC-TGGCCAGAAT----------
VoleM.a. TTCTCCAGGCCTGC----TGTGGGCTTAGTGCTGACAC-TGGCCAGAGT----------
VoleM.t. TTCTCCAGGCCTGC----TGTGGGCTTAGTGCTGACAC-TGGCCAGAGT----------
Mouse TTCTCTAAGCCTAC----TG-GGTATAAGTGGTGACTT-TGGCCAGAGTCATAGTGGATC
CMXIST2/CMXISTREV2
Cow GCAGATGATCGTTGGCCAACTGTTGGCAGAAGAGGAATCCCTGCCTTCCTCAAGAGGAAC
Horse GCAGATGATCGTTGGCCAACAGGTGGCAGAAGAGGAATCCCTGTCTTCCTCAAGAGGAAC
Coastmole GCAGATGATCGTTGACCA-CAGGTGGCAGAAGAGGAATCCTTGCCTTTCTCAAGAGGAAC
Human ACAGATGATCGTTGGCCAACAGGTGGCAGAAGAGGAATTCCTGCCTTCCTCAAGAGGAAC
Rabbit ACAGATGATCG----------AAGGGAAATCC-TGCCTTTCTCAGGAGGAAC
VoleM.r. --AAA---GCCTTGGCCA----------GAGGTGGAATCC-TGCCTCCTTCCACAGGCTC
VoleM.a. --AAA---GCCCTGGCCA----------GAGGTGGAATCC-TGCCTCCTTCCACAGGCTC
VoleM.t. --AAA---GCCCTGGCCA----------GAGGTGGAATCC-TGCCTCCTTCCACAGGCTC
Mouse ACAAA---TCACTGGTGAA---------GAGGTAGAATCC-TACCTTCTTCCAAAA----
SKXIST2
Cow ACCTACCCCGTGGCTAATGCTGGGGTCGG----------
Horse ACCTACCCCGTGGCTAATGCTGGGGTCGG----------
Coastmole ATCTACCCCGTGGCTAATGCTGGGGTCGGGATATCTACCCNTGGCTATGGCTGGGGTCGG
Human ACCTACCCCTTGGCTAATGCTGGGGTCGG----------
Rabbit ATCTACCCCGTGGCTAATGCTGGGGTCGG----------
VoleM.r. GTCTACCCCTTGGCTAATGCTGGGGTTGA----------
VoleM.a. GTCTACCCCATGGCTAATGCTGGGGTTGA----------
VoleM.t. GTCTACCCCATGGCTAATGCTGGGGTTGA----------
Mouse -TCTACCCCATGACTATTGCTGGGGTTGC----------
Cow ATTTT----------TGATA----------
Horse ATTTTGATTTCTATTT----CTTGGATGTCAGTCATATATAGTTTGATTATGTGGTTTGC
Coastmole ATTTTGATTTCTATTTATTTCTTGGATGTCAGTCATATATAGTTTGATTGTGTGGTTTGC
Human ATTTTGATTTATATTTATCTTTTGGATGTCAGTCATACA--GTCTGATTTTGTGGTTTGC
Rabbit ATTTTGATTTATATTATTTTTTTGGATGTCAGTCTTA-ATAGTCTGACTATGTGGTTTGC
VoleM.r. AATTTCATTTATGTGAAGGTTTTGGATGTCAGTGAA-----GTTTGATAGTGTGGTTTGC
VoleM.a. AATTTCATTTATGTGAAGGTTTTGGATGTCAATGAA-----GTTTGATAGTATGGTTTGC
VoleM.t. AATTTGATTTATGTGAAGGTTTTGGATGTAAGTGAA-----GTTTGATAGTGTGGTTTGC
Mouse ATTTTGATTTCAATGAATATTTTGGATGCCAACGACAC---GTCTGATAGTGTGCTTTGC
Cow ----------
Horse TAGTGTTCGA-TTTAAGCCTTAAGTGACTAT-ATGGAAATGTATTTAGGGGCTTTATTT-
Coastmole AAGTATTCTA-CTTAAACCTTAAGTGACTGA-ATGGAAATGTATTTAGGGACTTTGTTTA
Human TAGTGTTTGAATTTAAGTCTTAAGTGACTATTATAGAAATGTATTAAGAGGCTTTATTT-
Rabbit TAGTGTTCCAATTTAAGTCTTAAGTGACTAGTATAGAAATGTATTTTGTGACTTTTTTAA
VoleM.r. TAGTGCTTGAATCTAAAGCTGGAGTGATTATA---AAAATGTATTTAGGAGCTTTACTG-
VoleM.a. TAGTGCTTGAATCTAAAGCTGGAGTGATTATA---AAAATGTATTTAGGAGCTTTACTG-
VoleM.t. TAGTGCTTGAATCTAAAGCTGGAGTGATTATA---AAAATGTATTTAGGAGCTTTACTG-
Mouse TAGTGTTTGAATTTAAAACCGAAGTGATTGTTTTCAAAATGTATTTACGGATTTGCTTA-
CMXISTREV3
Cow ----------
Horse ---GTAGAATTC----------
Coastmole TCTGTAGGATTCATTTTAAGGCCAATTAATCAGTTCAT-TTTCGGTGTCCCTAAAATTTC
Human ---GTAGAATTCACTTTAATTACATTTAATGAGTTTTTGTTTTGAGTTCCTTAAAATTCC
Rabbit TTTGTAGAATTCATTTTAGTCGCACTTAATGAGCTCTCATTGCGCATTCCATAGAATCCC
VoleM.r. CTTGTAGAATTAATTTGAATCACC----GTATGCTTTCACTCCAGAGTCCTTAGTGTTT-
VoleM.a. CTTGTAGAATTCAATTGAATCACC----GTATGCTTTCACTCCAGAGTCCTTAGTGTTT-
VoleM.t. CTTGTAGAATTCAATTGAATCACC----GTATGCTTTCACTCCAGAGTCCTTAGTGTTT-
Mouse CTTGTTGAATTCATTTTAATTACCTTTAGTGAATTGTTACTTTGGAGTCCTTAAAGTTT-

```
Cow          ------------------------------------------------------------
Horse        ------------------------------------------------------------
Coastmole    TTAATGTTTTCCATTTCTTTTTTGTAAATTGCTTAAAAGCACTTTGGCTT-TAGTTATTT
Human        TTAAAGTTTTTAGCTTCTCATT-ACAAATTCCTTAACCTTTTTTTGGCAG-TAGATAGTC
Rabbit       TTAAATTTCTCTAT------------AAATACATAAACTAGTTGTGGCAGGCAGATAGCC
VoleM.r.     TTAGCTTCTTTAA-------------AATTCATTAA--CTTTTTTGGCAGAT-GATACTC
VoleM.a.     TTAGCTTCTTTAA-------------AATTCATTAA--CTTTTTTGGCAGAT-GATACTC
VoleM.t.     TTAGCTTCTTTAA-------------AATTCATTAA--CTTTTTTGGCAGAT-GATACTC
Mouse        TCAA----------------------------TAA--TTTTTTTGGCAGAT-GATACTC

Cow          ------------------------------------------------
Horse        ------------------------------------------------
Coastmole    A--GT------CAAATAATTT--GACATTTAA-----------TAAGCT
Human        AAAGT------CAAATCATTTCTAATGTTTTA-----------AAAATG
Rabbit       AAATC------CACCTCATCT--GACATTTAA-----------AAACTT
VoleM.r.     AAGT-------CACTTAAATTT--ACGTTCTT-----------TCAAACT
VoleM.a.     AAGT-------CACTTAAATTT--ACGTTCTT-----------TCAAACT
VoleM.t.     AAGT-------CACTTAAATTT--ACGTTCTTCGTTCTTTCAATCAAACT
Mouse        AAATTACTTGGCACTTAAATGT--ACTTTCTT-----------TCAAACT
```

PERCENT SEQUENCE SIMILARITY

	Coast mole	Cow	Horse	Human	Rabbit	Vole M.r.	Vole M.a.	Vole M.t.
Cow	75							
Horse	77	84						
Human	70	76	77					
Rabbit	71	74	77	77				
Vole M.r.	12	16	21	20	24			
Vole M.a.	12	16	21	19	23	98		
Vole M.t.	12	24	21	19	24	98	98	
Mouse	11	16	21	17	22	68	69	68

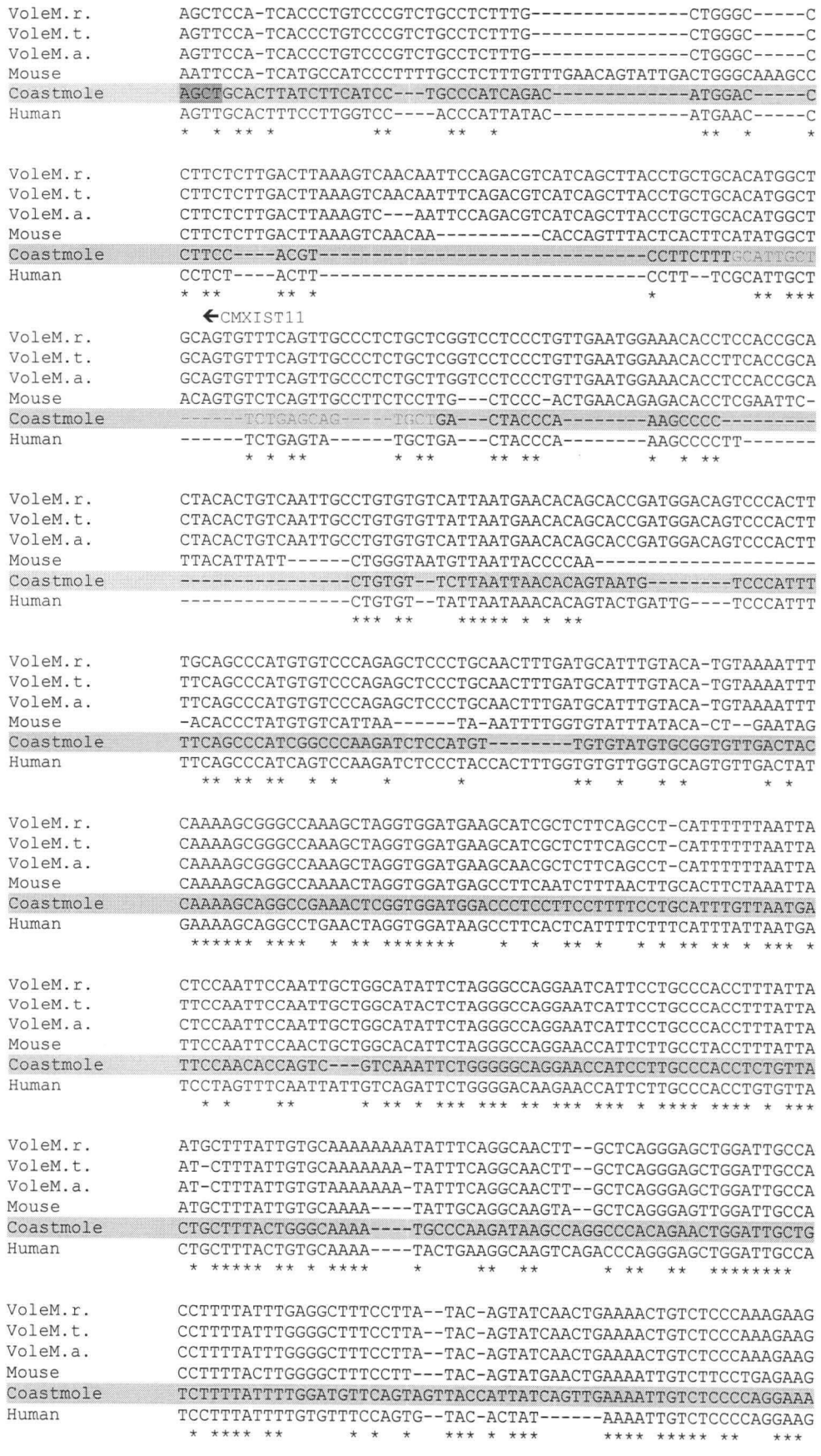
Figure 12a: Aligned sequences in the 5'region of exon one.
The Coast mole *Xist* fragment corresponding to the 5' region of exon one is aligned with the Xist seqeunces from human (M97168; bases 1074-2519), cow(AF104906:1102-2182), rabbit (U50910;1899-3278), horse (U50911;4810-6055), mouse (L04961;2029-2237) and the three vole species, *M. arvalis* (AJ310129;1031-2507), *M. rossiameridionalis* (AJ310130;1041-2534), and *M. transcaspicus* (AJ310127:1034-2538). Primers designed from conserved sequence are green, mole-specific primers are red and the two *AluI* sites are highlighted in turquoise.

```
Rabbit          TGCTTGTTAGCAGCTTTGCATGAATTGTGGTACTGGGATGCCGGAAAACRAGATTGTCCT
Human           GGCTTGTTAGGAGCTTTGCGTGA-TTGTTGTATCGGGAGGCAGTAAGA----ATCATCTT
Horse           GGCTCGTTAGGAGCTTTGTATGAATTAGGGTATCGTGAGGTGGAGAAACAGGATGGTCCT
Coastmole       AGCTTGTTAGGAACGCTGCAGTTACTCTAGTATCGCGAGGTCTCGAAACGGGGTTGTGCT
                 *** ***** * *  **       *   ***  * ** *     * *     *  *  *

Rabbit          GCAGCAG--CACGAGGCGTTAACTTGAAAATG-------AGACAGGACTTAAAAGTGCTG
Human           TTATCAG--TACAAGGGACTAG-TTAAAAATGGAAGGTTAGGAAAGACT--AAGGTGCAG
Horse           GTGTTATGTTAAAAGAGGCTAA-TCGAAAATGGGAG---GGAGGAAACTT--AGGCACAG
Coastmole       GT--CAC--TACCAGAGGCACG-TTGAGAATGACAA---GGAGGGGAGTT-AAGGCACTG
                    *    *  **       *  * ****        *     * *   * *  * *

Rabbit          GGATCAAAATGGCGCTTTTGCCATTGCAGCGTTTCTCAGCATGGCTCCCTGTGCTTTGTT
Human           GGCTTAAAATGGCGATTTTGACATTGCGGCATTGCTCAGCATGGCGGGCTGTGCTTTGTT
Horse           GGTCCGAAATGGCGATTTTGACTTTGCGACATTGCTCAGCATGGCTGCCTG--CTTTGTT
Coastmole       ACATCAAAATGGCGGCGTTAACTTTGTGGCGTTGCTCAGCATGGCTGCCTGT--TTTGTG
                      ********   **  * ***   * ** ***********   ***   *****

Rabbit          AAATCGTCTAAAATGGCGGATTCAGTTCGG-CCGCAACGTTAAAGTGGCGGGAAGCCCAC
Human           AGGTTGTCCAAAATGGCGGATCCAGTTCTG-TCGCAGTGTTCAAGTGGCGGGAAGGCCAC
Horse           AG--TGTCCAAAATGGCGGACACAGTTTTG-CCGCAGTGTTCAAGTGGCGGGAAGGCCAC
Coastmole       AA--TGTCCAAAATGGCGGATTCATTTTTTTCCGCAGTGTCCAAGTGGCGGGAAGGCCAC
                *    *** ***********  ** **     ****  **  ************* ****

Rabbit          ATCATGATGGGTGGGGTCTTTTTTTGTAATGCGCAGCATGATGGTGRCAGAGTTCGGTCA
Human           ATCATGATGGGCGAGGCTTTGTTAAGTGGT---TAGCATGGTGGTGGACATGTGCGGTCA
Horse           ATCATGGTGG---TGTCTTTGTTC--TAGTGCGCAGCATGGCGGTGGAAATATTCTGTTA
Coastmole       ATCCTGGTGG---CGTCTTTGTTC--TAATGTGCAGCATGGCGGTGGAAATATTCTGTTA
                *** ** ***    *   ** **   *  *    ******  ****      * * ** *

Rabbit          CACAGTGAAAGATGGCGGCTGAAACCCTTGCCGCAGTGTAAAATATGGCGGGCGTCTTTG
Human           CACAGGAAAAGATGGCGGCTGAAGGTCTTGCCGCAGTGTAAAACATGGCGGGCCTCTTTG
Horse           CATGGCAAAAGATGGCGGCTGAAGTACTTGCCGCAGTATAAAATATGGCGGGC--TTTTG
Coastmole       CATGCTAAAAGATGGCGGCTAGAGCACTTG------TGTAA--CATGGCGGCC---TTTG
                **     *************  *   ****      * ***   ******* *   ****

Rabbit          TCTTTGCCGTGTGCATTTCGTGGCGGGTTTTGCCGCAGGGACAATATGGCTGACGATATC
Human           TCTTTGCTGTGTGCTTTTCGTGTTGGGTTTTGCCGCAGGGACAATATGGCAGGCGTTGTC
Horse           TGTTTGCCGTGTGCATTTCCTGACAGGTTTTGCCGCAGGGACAATATGGCAGACCTTGTC
Coastmole       TCTTTGCTGTGTGTGTTTCCTGGCAGGTTTTGCCTCGGGGATAGTATGGCAGGCCTTATT
                * ***** *****  **** **   ********* * **** * ****** * *  * *

Rabbit          ATGTGGTTATCATGGAGATTTGTCACGTGGACGTCATGGCGCGT-----------TTTT
Human           ATATGTATATCATGGCT-TTTGTCACGTGGACATCATGGCG-------------------
Horse           ATGTGGATATCATGGCAGTTCGTCACGTGGACGTCATGGCGGGGGG---------ATTTT
Coastmole       ATGTGGAAGTCTTGGCAATTTGTCACGTGGATGTCTTGGCAGAGGGTCCACAGGGGCGTT
                ** **    ** ***   ** **********  ** ****

Rabbit          GCTCGGTAAATT---GGCGGGCTTTGCACAAGGTGGACCTACGGCATTGTTAAAGATGGC
Human           ----------------------------------GGCTTGCCGCATTGTTAAAGATGGC
Horse           GCCCGGTACATTCTTGGCGGGCTTTGCACCAGGTGGGCCTGCCGCATTGTTAAAGATGGC
Coastmole       CGTCTGCATGTTCTTGGCGGGCTCAGCACCTGGTGGTATTGCCGCATTGTTAAAGATGGC
                                                 *   * * *****************

Rabbit          GGG----------CGTTGCCGCATTTTGCCGGTC-------------------AAAGCA-
Human           GGG----------TTTTGCCGCCTAGTGCCACGC-------------------AGAGCG-
Horse           GGGGTTTGGCGCGTTTTGGCGCGTTTTGCCGCGTTTGCTGCGTTTTGCCCCGAAAAGTG-
Coastmole       GGG----------CTTTGCCGCGTT-TGTGGCGG-------------------AAAGCNC
                ***            *** *** *  **                          * **

Rabbit          ---GGAGAAAAC------CGG---GTGCTGGATTGCCGCATGACTTAACCAATCAGAAAT
Human           ---GGAGAAAAGGTGGGATGGACAGTGCTGGATTGCTGCATAACCCAACCAATTAGAAAT
Horse           ---GAAGAAAGG-TGGGGTGGACAGTGTTGGACTGCCGCACGACCCACCCAATCGAGAAG
Coastmole       GAAGGAGAAACA-TGGGGTGGGGAGTGNTCAATTGCTGCTTTACATACCCAATCAGAGA-
                   * *****          **   *** *  * *** **   **  * *****     *

Rabbit          GAG-GGTGGA-TTAGACACAACCAATTAGCATGGAGGATGGAATTAGACTGTCGATGCAC
Human           GGG-GGTGGAATTGATCACAGCCAATTAGAGCAGAAGATGGAATTAGACTGATGACACAC
Horse           GGGTGGTAGAATTGGGAACAACCAATTAG--TAGAGGATGGAATTAGACTA---------
Coastmole       GCATGTTAGAAAAGAGTATGACCAATTAG--TGGAAGATGGAATTACANTGG--------
                *   * * **       *   ********    ** ********** * *
```

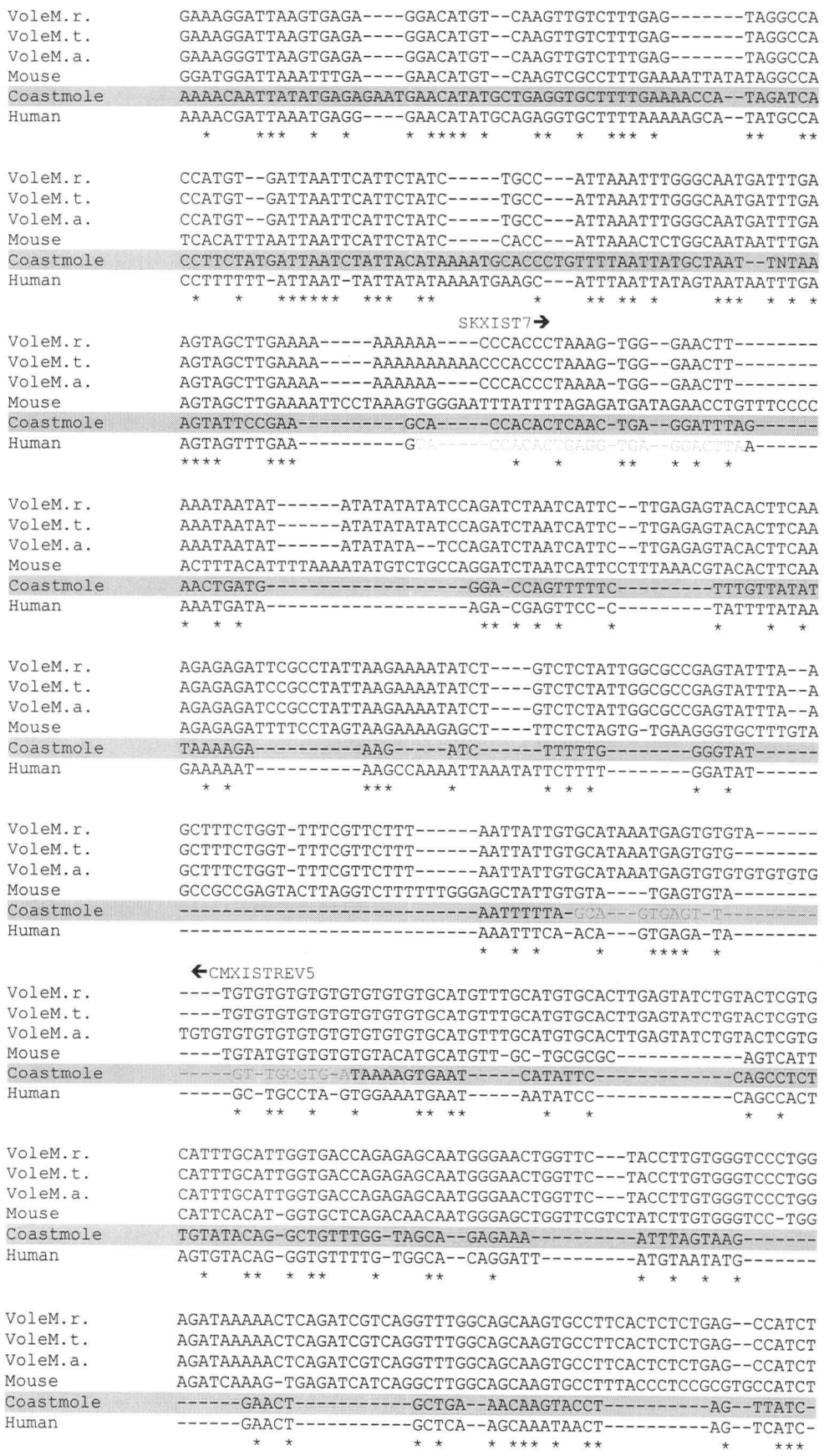
```
Rabbit          TGTCCAGGTACTGAGCATGGTCATGTGGGCTGAGTTAGCATGGCTCTTTGCAGCTTTCTT
Human           TGTCCAGCTACTCAGCGAAG--ACCTGGG-TGAATTAGCATGGCACTTCGCAGCTGTCTT
Horse           -----------------------------TGAGGTAGTGTAGCACTGCGCTGTGGTCTT
Coastmole       -----------------------------TGAGTTAGCATGGCTCCTCACTGAAGTCTT
                                              ***  ***  * ** *    * *   ****

Rabbit          TTTTCAGCCAGTCAGGACGAAGAAGTGGAGCGGCCATGTA-GTGTGCCC-CTCCCAGTGG
Human           T----AGCCAGTCAGGAGAAAGAAGTGGAGGGGCCACGT--GTATGTC---TCCCAGTGG
Horse           TATTCAGCCAGTCAGGTGGAGGAAGCAGAGAGGTCACGTCTCTACGTCGGCTCCCAGTGG
Coastmole       TTTTCAGCCAGTGAGAAGGCAGTA-TGGAGCAGCCACGTTTCTAAGTCGGCTCCCCGTGG
                *    ******* **      * *   ***  * ** **   *  * *   **** ****

Rabbit          GTGGTATTAAAGTTCTTTCCAAGGTCTTTTCAAGGACATTTGGCCTTTCTACCTCCCTTC
Human           GCGGTACACCAGGTGTTTTCAAGGTCTTTTCAAGGACATTTAGCCTTTCCACCTCTGTCC
Horse           GTGGTAA-----------CCAAGGTCTTTCCAAGGACATCTGGCCTTTCCACCTCCCTCC
Coastmole       GTGGTGC-----------CCAAGGTCCTTCCAAGGACATGTGGCCTTTCCACCTCCCTCC
                * ***             ******* ** ********* * ******* *****  * *

Rabbit          CCTCTTGTTTGTCCTCTCCTCTCCAGACATGGTCTCTTGCAGTGCTGGATATCTGGCTAT
Human           CCTCTTATTTGTCCCCTCCTGTCCAG-TGCTGCCTCTTGCAGTGCTGGATATCTGGCTGT
Horse           CCTCTCACTTGCTCCCTCCCCTCCGG-CATTGACTCTTGCAGTGCTGGATAGCTGGCTGT
Coastmole       CTCCTTCCCTGCTCCCTCCGCTCCAG-CATTACCTCTTGCACTGCTGGGT-TTTGACTGT
                *  **    **  * ****  *** *       ******** ****** *   ** ** *

Rabbit          GTGGGCTGAACCCCACCCACATTCCTCTGTATTGATGCAATGGCCTTTGGCTAA------
Human           GTGGTCTGAACCTCCCTC-CATTCCTCTGTATTGGTGCC-TCACCTAAGGCTAAGTATAC
Horse           GCGGGCTGAACCTCACCC-CATTCCTCTGCGTTGCTGCTGTGGCCTCAGGCTAG------
Coastmole       GCTGGCTGAACCT-ACAC-AATTCCTCTTCATTGCTGAAGTGGCCTGAGGTTAGACTTA-
                *  * *******   * *  ********   *** **   *  ***  ** **

Rabbit          -------------------------------TCTACTCCC---GCCCTGG----------
Human           CTCCCCCCCCACCCCCCAACCCCCCCAACTCCCCACCCCC---ACCCCCCACCCCCCACC
Horse           -------------------------------TCTACTCCC----CCCTGG----------
Coastmole       -------------------------------CCTGGCCCCCTTGCCCCAGG---------
                                               *    ***    ***

Rabbit          ----------------------------------CCTGTGGTCCTCCCTGC-----ACTG
Human           TCCCCACCCCCCTACCCCCCTACCCCCCTACCCCCCTCTGGTCTGCCCTGCACTGCACTG
Horse           ----------------------------------CTTCTGCTCTTCC--GC-----ACTG
Coastmole       ----------------------------------CCCCGCTCTTCC--AC-----TCTG
                                                   * **  **   *      ***

Rabbit          TGGCCACCGG--CAGTGCTCCAGGCCTGCCTGGTGTGGACATGGTGGTGAGCTGTGGCAA
Human           TTGCCATGGG--CAGTGCTCCAGGCCTGCTTGGTGTGGACATGGTGGTGAGCCGTGGCAA
Horse           TGGCCAGGGG--CAGTGCTCCATGCCTGCCAAGTATGAATATGGCGGTGAGTCGTGGCCA
Coastmole       TGGCCCCCGGAGCAGTGCTCCATGCTGGCCAAGTGTGAACATGGTGATCAGTCGAGGCAA
                * ***   **  ********** **  **   ** ** * **** * * **  * *** *

Rabbit          GGACCAGAATGGATCACAGATGATCG-------------------AAGGGAAATCCTGCC
Human           GGACCAGAATGGATCACAGATGATCGTTGGCCAACAGGTGGCAGAAGAGGAATTCCTGCC
Horse           AGACCAGAATGGATCGCAGATGATCGTTGGCCAACAGGTGGCAGAAGAGGAATCCCTGTC
Coastmole       GGACCAGAATGGATCGCAGATGATCGTTGACCA-CAGGTGGCAGAAGAGGAATCCTTGCC
                 ************** **********                    *  ****  * ** *

Rabbit          TTTCTCAGGAGGAACATCTACCCCGTGGCTAATGCTGGGGTCGG-----------------
Human           TTCCTCAAGAGGAACACCTACCCCTTGGCTAATGCTGGGGTCGG-----------------
Horse           TTCCTCAAGAGGAACACCTACCCCGTGGCTAATGCTGGGGTCGG-----------------
Coastmole       TTTCTCAAGAGGAACATCTACCCCGTGGCTAATGCTGGGGTCGGGATATCTACCCNTGGC
                ** **** ******** ******* ******************

Rabbit          ---------------ATTTTGATTTATATTATTTTTTTGGATGTCAGTCTTAA-TAGTCT
Human           ---------------ATTTTGATTTATATTTATCTTTTGGATGTCAGTCATA--CAGTCT
Horse           ---------------ATTTTGATTTCTATT----TCTTGGATGTCAGTCATATATAGTTT
Coastmole       TATGGCTGGGGTCGGATTTTGATTTCTATTTATTTCTTGGATGTCAGTCATATATAGTTT
                           ********** ****    * ************ **   *** *

Rabbit          GACTATGTGGTTTGCTAGTGTTCCAATTTAAGTCTTAAGTGACTAGTATAGAAATGTATT
Human           GATTTTGTGGTTTGCTAGTGTTTGAATTTAAGTCTTAAGTGACTATTATAGAAATGTATT
Horse           GATTATGTGGTTTGCTAGTGTTCGA-TTTAAGCCTTAAGTGACTA-TATGGAAATGTATT
Coastmole       GATTGTGTGGTTTGCAAGTATTCTA-CTTAAACCTTAAGTGACTG-AATGGAAATGTATT
                ** * ********** *** **  *  ****  ***********   ** **********
```

```
Rabbit          TTGTGACTTTTTTAATTTGTAGAATTCATTTTAGTCGCACTTAATGAGCTCTCATTGCGC
Human           AAGAGGCTTT----ATTTGTAGAATTCACTTTAATTACATTTAATGAGTTTTTGTTTTGA
Horse           TAGGGGCTTT----ATTTGTAGAATTC--------------------------------
Coastmole       TAGGGACTTTGTTTATCTGTAGGATTCATTTTAAGGCCAATTAATGAGTTCAT-TTTCGG
                  * * ****    ** ***** ****

Rabbit          ATTCCATAGAATCCCTTAAA--------TTTCTC---TATAAATACATAAAC-TAGTTGT
Human           GTTCCTTAAAATTCCTTAAAGTTTTTAGCTTCTCAT-TACAAATTCCTTAACCTTTTTTT
Horse           ------------------------------------------------------------
Coastmole       TGTCCCTAAAATTTCTTAATGTTTTCCATTTCTTTTTTGTAAATTGCTTAAAAGCACTTT

Rabbit          GGCAGGCAGATAGCCAAATCCACCTCA--TCTGACATTTAAAAACTT
Human           GGCAG-TAGATAGTCAAAGTCAAATCATTTCTAATGTTTTAAAAATG
Horse           ----------------------------------------------
Coastmole       GGCTT-TAGTTATTTAG--TCAAATAA--TTTGACATTTAATAAGCT
```

Figure 12b: Aligned sequences in the 5'region of exon one (without rodents, cow)
The Coast mole Xist fragment corresponding to the 5' region of exon one is aligned with the human, equine and lepine sequences. The rodent sequences were omitted so that the high level of homology among the genes of the remaining species could be visualized. Although the bovine gene also shows high homology to the sequences in this figure, it was not included due to its small size. The locations of the various primers and the sequence similarities are shown in Part A of this figure.


VoleM.r. AGCTCCA-TCACCCTGTCCCGTCTGCCTCTTTG----------CTGGGC-----C
VoleM.t. AGTTCCA-TCACCCTGTCCCGTCTGCCTCTTTG----------CTGGGC-----C
VoleM.a. AGTTCCA-TCACCCTGTCCCGTCTGCCTCTTTG----------CTGGGC-----C
Mouse AATTCCA-TCATGCCATCCCTTTTGCCTCTTTGTTTGAACAGTATTGACTGGGCAAAGCC
Coastmole AGCTGCACTTATCTTCATCC---TGCCCATCAGAC----------ATGGAC-----C
Human AGTTGCACTTTCCTTGGTCC----ACCCATTATAC----------ATGAAC-----C
VoleM.r. CTTCTCTTGACTTAAAGTCAACAATTCCAGACGTCATCAGCTTACCTGCTGCACATGGCT
VoleM.t. CTTCTCTTGACTTAAAGTCAACAATTTCAGACGTCATCAGCTTACCTGCTGCACATGGCT
VoleM.a. CTTCTCTTGACTTAAAGTC---AATTCCAGACGTCATCAGCTTACCTGCTGCACATGGCT
Mouse CTTCTCTTGACTTAAAGTCAACAA----------CACCAGTTTACTCACTTCATATGGCT
Coastmole CTTCC----ACGT----------CCTTCTTTGCATTGCT
Human CCTCT----ACTT----------CCTT--TCGCATTGCT
CMXIST11
VoleM.r. GCAGTGTTTCAGTTGCCCTCTGCTCGGTCCTCCCTGTTGAATGGAAACACCTCCACCGCA
VoleM.t. GCAGTGTTTCAGTTGCCCTCTGCTCGGTCCTCCCTGTTGAATGGAAACACCTTCACCGCA
VoleM.a. GCAGTGTTTCAGTTGCCCTCTGCTTGGTCCTCCCTGTTGAATGGAAACACCTCCACCGCA
Mouse ACAGTGTCTCAGTTGCCTTCTCCTTG---CTCCC-ACTGAACAGAGACACCTCGAATTC-
Coastmole ------TCTGAGCAG-----TGCTGA---CTACCCA--------AAGCCCC---------
Human ------TCTGAGTA------TGCTGA---CTACCCA--------AAGCCCCTT-------
VoleM.r. CTACACTGTCAATTGCCTGTGTGTCATTAATGAACACAGCACCGATGGACAGTCCCACTT
VoleM.t. CTACACTGTCAATTGCCTGTGTGTTATTAATGAACACAGCACCGATGGACAGTCCCACTT
VoleM.a. CTACACTGTCAATTGCCTGTGTGTCATTAATGAACACAGCACCGATGGACAGTCCCACTT
Mouse TTACATTATT------CTGGGTAATGTTAATTACCCCAA----------
Coastmole ----------CTGTGT--TCTTAATTAACACAGTAATG--------TCCCATTT
Human ----------CTGTGT--TATTAATAAACACAGTACTGATTG----TCCCATTT
VoleM.r. TGCAGCCCATGTGTCCCAGAGCTCCCTGCAACTTTGATGCATTTGTACA-TGTAAAATTT
VoleM.t. TTCAGCCCATGTGTCCCAGAGCTCCCTGCAACTTTGATGCATTTGTACA-TGTAAAATTT
VoleM.a. TTCAGCCCATGTGTCCCAGAGCTCCCTGCAACTTTGATGCATTTGTACA-TGTAAAATTT
Mouse -ACACCCTATGTGTCATTAA------TA-AATTTTGGTGTATTTATACA-CT--GAATAG
Coastmole TTCAGCCCATCGGCCCAAGATCTCCATGT--------TGTGTATGTGCGGTGTTGACTAC
Human TTCAGCCCATCAGTCCAAGATCTCCCTACCACTTTGGTGTGTTGGTGCAGTGTTGACTAT
VoleM.r. CAAAAGCGGGCCAAAGCTAGGTGGATGAAGCATCGCTCTTCAGCCT-CATTTTTTAATTA
VoleM.t. CAAAAGCGGGCCAAAGCTAGGTGGATGAAGCATCGCTCTTCAGCCT-CATTTTTTAATTA
VoleM.a. CAAAAGCGGGCCAAAGCTAGGTGGATGAAGCAACGCTCTTCAGCCT-CATTTTTTAATTA
Mouse CAAAAGCAGGCCAAAACTAGGTGGATGAGCCTTCAATCTTTAACTTGCACTTCTAAATTA
Coastmole CAAAAGCAGGCCGAAACTCGGTGGATGGACCCTCCTTCCTTTTCCTGCATTTGTTAATGA
Human GAAAAGCAGGCCTGAACTAGGTGGATAAGCCTTCACTCATTTTCTTTCATTTATTAATGA
VoleM.r. CTCCAATTCCAATTGCTGGCATATTCTAGGGCCAGGAATCATTCCTGCCCACCTTTATTA
VoleM.t. TTCCAATTCCAATTGCTGGCATACTCTAGGGCCAGGAATCATTCCTGCCCACCTTTATTA
VoleM.a. CTCCAATTCCAATTGCTGGCATATTCTAGGGCCAGGAATCATTCCTGCCCACCTTTATTA
Mouse TTCCAATTCCAACTGCTGGCACATTCTAGGGCCAGGAACCATTCTTGCCTACCTTTATTA
Coastmole TTCCAACACCAGTC---GTCAAATTCTGGGGGCAGGAACCATCCTTGCCCACCTCTGTTA
Human TCCTAGTTTCAATTATTGTCAGATTCTGGGGACAAGAACCATTCTTGCCCACCTGTGTTA
VoleM.r. ATGCTTTATTGTGCAAAAAAAATATTTCAGGCAACTT--GCTCAGGGAGCTGGATTGCCA
VoleM.t. AT-CTTTATTGTGCAAAAAAA-TATTTCAGGCAACTT--GCTCAGGGAGCTGGATTGCCA
VoleM.a. AT-CTTTATTGTGTAAAAAAA-TATTTCAGGCAACTT--GCTCAGGGAGCTGGATTGCCA
Mouse ATGCTTTATTGTGCAAAA----TATTGCAGGCAAGTA--GCTCAGGGAGTTGGATTGCCA
Coastmole CTGCTTTACTGGGCAAAA----TGCCCAAGATAAGCCAGGCCCACAGAACTGGATTGCTG
Human CTGCTTTACTGTGCAAAA----TACTGAAGGCAAGTCAGACCCAGGGAGCTGGATTGCCA
VoleM.r. CCTTTTATTTGAGGCTTTCCTTA--TAC-AGTATCAACTGAAAACTGTCTCCCAAAGAAG
VoleM.t. CCTTTTATTTGGGGCTTTCCTTA--TAC-AGTATCAACTGAAAACTGTCTCCCAAAGAAG
VoleM.a. CCTTTTATTTGGGGCTTTCCTTA--TAC-AGTATCAACTGAAAACTGTCTCCCAAAGAAG
Mouse CCTTTTACTTGGGGCTTTCCTT---TAC-AGTATGAACTGAAAATTGTCTTCCTGAGAAG
Coastmole TCTTTTATTTTGGATGTTCAGTAGTTACCATTATCAGTTGAAAATTGTCTCCCCAGGAAA
Human TCCTTTATTTTGTGTTTCCAGTG--TAC-ACTAT------AAAATTGTCTCCCCAGGAAG

```
VoleM.r.        GAGGATGAGCTAGTACTTT-CTTTACACCCTTCCTCCAGGA----GCTGACCGCCTGCTT
VoleM.t.        GAGGATGAGCTAGTACTTT-CTTTACACCCTTCCTCCAGGA----GCTGACCGCCTGCTT
VoleM.a.        GAGGATGAGCTAGTACTTT-CTTTACACCCTTCCTCCAGGA----GCTGACCGCCTGCTT
Mouse           GAAGCT----TAGCACTTTTCTTTCCATTCTTCCTCCAGGAAGGAGCCAACTGTCTGCTT
Coastmole       GAAG-T----TAGCGTTGT-CTCT----------TTCCTGA----GCAGAGTGCCTGGCT
Human           GAAGGT----TGGCACTTT-CTCTGCATTCTTCTTTCCAGA----GCAGATTGCCTGGTT
                ** * *    * *   * * ** *          * *  **    **  *  * ***  *

VoleM.r.        AACAAACTC--AACTTGTTTCTC--TTGGTATATTGCTAC-GTACAGTGCCAACTGCCAG
VoleM.t.        AACAAACTC--AACTTGTTTCTC--TTGGTATATTGCTAC-GTACAGTGCCAACTGCCAG
VoleM.a.        AACAAACTC--AACTTGTTTCTC--TTGGTATATTGCTAC-GTACAGTGCCAACTGCCAG
Mouse           AAGAAACTTTAAGCCCGATTTT------GTATATTGCTACTGTACAGGACCAACTGCCAG
Coastmole       GTAAATCTC-----TCATC-------TTATATATTGCTATTGTAAAGTGCCAGTTGCCAG
Human           AAGAATCTC-----TTGTTGTCCCTTCTGTATATTGTTATTGTAAAGTGCCAAATGCCAG
                   ** **                    ******* **  *** **  ***  ******

VoleM.r.        A-----------AAAGTTATTGATAGTTTAT--------GCCTTTAAGAAAGGCCTCTGG
VoleM.t.        A-----------AAAGTTATTGATAGTTTAT--------GCCTTTAAGAAAGGCCTCTGG
VoleM.a.        A-----------AAAGTTATTGATAGTTTAT--------GCCTTTAAGAAAGGCCTCTGG
Mouse           A-----------AAAGTTATTGATAATTTTA--------TTCCTTAAGAAAGGCATTTGG
Coastmole       AG-GCAACCA--AAAGTTGCTATTAATTTTTTA-----TTTTCTAAAGAAAGGTATCTGG
Human           GATACAGCCAGAAAAATTGCTTATTATTATTAAAAAAATTTTTTTAAGAAAGACATCTGG
                           *** **  *  *  **              * *******   * ***
                        CMXIST11➔
VoleM.r.        CTTGCCAGGTGGAATTGATTATGTGATCATTAGCTTTTCTGAAGTAAAAATATCTTGCTT
VoleM.t.        CTTGCCAGGTGGAATTGATTATGTGATCATTAGCTTTTCTGAAGTAAAAATATCTTGCTT
VoleM.a.        CTTGCCAGGTGGAATTGATTATGTGATCATTAGCTTTTCTGAAGTAAAAATATCTTGCTT
Mouse           ATTGCAAGGTGGA-TTGACTGTGAGATCATTAGCTTTTGTGAAGTAAAAATAGCC-ATTT
Coastmole       AATATCAGGTGGAGTTGATAACGTGGTCATTAAATATT-TGAAGTCCTAAAACCC-ATTT
Human           ATTGTAGGGTGGACTCGATAACCTGGTCATTATTTTTT-TGAAGCCAAAATATCC-ATTT
                  *    ****** * **      * ******  * ** *****    ** * *    **

VoleM.r.        CTACTATATATCTGATGACCAAGTAGTGTCTCTCATTTTA-CTGAGGGTGGCGAGTCTGT
VoleM.t.        CTACTATATATCTGATGACCAAGCAGTGTCTCTCATTTTA-CTGAGGGTGGCGAGTCTGT
VoleM.a.        CTACTATATATCTGATGACCAAGCAGTGTCTCTCATTTTA-CTGAGGGTGGCGAGTCTGT
Mouse           GTGTCATGTTTCTGAAGACTAAGCAGTGTCTC--AGTGTA-CTGAGGGTGATGAGTCTGT
Coastmole       ATATCATGTACCTAATGAC----CAGTGTCTCTCATTTTA-CTAAGGGTGGTGGGTCTGN
Human           ATACTATGTACCTGGTGAC----CAGTGTCTCTCATTTTAACTGAGGGTGGTGGGTCTGT
                 *   ** *  **   ***     ********  * * ** ** ******  * *****

VoleM.r.        GGAGAAACCACTGTAACTGTTATTCTAATACCGTCAAAGTGGTTGTAGAATGGCCTGCTT
VoleM.t.        GGAGAAACCACTGTAACTGTTATTCTAATACCGTCAAAGTGGTTGTAGAATGGCCTGCTT
VoleM.a.        GGAGAAACCACTGTAACTGTTATTCTAATACCGTCAAAGTGGTTGTAGAATGGCCTGCTT
Mouse           GGAAAGATCAGTGCAACTATT-----------------------GCAGAAT-----GTTT
Coastmole       GGATAGACCACTGTGACTCTTGCT-----------------ACTTTATTAT----TGCTC
Human           GGATAGAACACTG--ACTCTTGCT-----------------ATTTTAATAT----CA---
                *** * * ** **  *** **                        *  **

VoleM.r.        AAGGCCACGATGTATGTTTGGTCTGTACTACCAGTTTAA----ATACAAAAGTCAATCTT
VoleM.t.        AAGGCCAGGATGTATGTTTGGTCTGTACTACCAGTTTAA----ATACAAAAGACAATCTT
VoleM.a.        AAGGCCAGGATGTATGTTTGGTCTGTACTACCAGTTTAA----ATACAAAAGTCAATCTT
Mouse           AAGACAAGTATCTTTGCTTGGTCTTTACTACAAGTTTAACAAAACGAAAAAGTCAATCTT
Coastmole       AAGTTAGTGA---GTTTTCAAGTGAAACT-----CTTAA------AGACCGGT--ACCTT
Human           AAGATA--------TTCTAGAGTGGAACT-----CTTAA-------GACCAGT--ATCTT
                ***           *  *        ***      ****        *   *   * ***

VoleM.r.        TGTGTGGCCTTCAGTATCCTTATCTTTAAGGAAAATGACC-AAACATTTTAATATTTAAA
VoleM.t.        TGTGTGGCCTTTAGTATCCTTATCTTTAAGGAAAATGACC-AAACATTTTAATATTTAAA
VoleM.a.        TGTGTGGCCTTTAGTATCCTTATCTTTAAGGAAAATGACC-AAACATTTTAATATTTAAA
Mouse           TGTGTGGCCTTTAGTATGATTAACTTTTTGGAAGATGACCTAAGCCTTCTAATCATTATA
Coastmole       TG----GGCTCTACCACCATTCATTTTTAGAAAAACTATGGAAACTTTATAGTCCTTAAA
Human           TGTGTGGGCTTTACCAGCATTCACTTTTAGAAAAACTACCTAAATTTTATAATCCTTTAA
                **    * **  *  *   **   ***  * ** *  *   **   ** ** *  **  *

VoleM.r.        TTTGGCCTGACATTGAACGCCAGCC---------CCTACTGAT-------GGGCTAC---
VoleM.t.        TTTGGCCTGACATTGAACGCCAGCC---------CCTACTGAT-------GGGCTGC---
VoleM.a.        TTTGGCCTGACATTGAACGCCAGCC---------CCTACTGAT-------GG-CTGC---
Mouse           TTTTGTCTGACATTGGTCACCAGTC---------CTTGCTTATTTTTTAAAAGGTGACT--
Coastmole       TTT--CTTCATCTGGAGCACCAACCAGTCCAGTCCTTATTTATTTCAATAAGATTGTTAT
Human           TTT--CTTCATCTGGAGCACCTGCC---------CCTACTTATTTCAAGAAGATTGCAGT
                ***    * *  * *  * **   *         * *  * **        *
```


VoleM.r. GAAAGGATTAAGTGAGA----GGACATGT--CAAGTTGTCTTTGAG-------TAGGCCA
VoleM.t. GAAAGGATTAAGTGAGA----GGACATGT--CAAGTTGTCTTTGAG-------TAGGCCA
VoleM.a. GAAAGGGTTAAGTGAGA----GGACATGT--CAAGTTGTCTTTGAG-------TAGGCCA
Mouse GGATGGATTAAATTTGA----GAACATGT--CAAGTCGCCTTTGAAAATTATATAGGCCA
Coastmole AAAACAATTATATGAGAGAATGAACATATGCTGAGGTGCTTTTGAAAACCA--TAGATCA
Human AAAACGATTAAATGAGG----GAACATATGCAGAGGTGCTTTTAAAAAGCA--TATGCCA
VoleM.r. CCATGT--GATTAATTCATTCTATC-----TGCC---ATTAAATTTGGGCAATGATTTGA
VoleM.t. CCATGT--GATTAATTCATTCTATC-----TGCC---ATTAAATTTGGGCAATGATTTGA
VoleM.a. CCATGT--GATTAATTCATTCTATC-----TGCC---ATTAAATTTGGGCAATGATTTGA
Mouse TCACATTTAATTAATTCATTCTATC-----CACC---ATTAAACTCTGGCAATAATTTGA
Coastmole CCTTCTATGATTAATCTATTACATAAAATGCACCCTGTTTTAATTATGCTAAT--TNTAA
Human CCTTTTTT-ATTAAT-TATTATATAAAATGAAGC---ATTTAATTATAGTAATAATTTGA
SKXIST7
VoleM.r. AGTAGCTTGAAAA-----AAAAAA----CCCACCCTAAAG-TGG--GAACTT--------
VoleM.t. AGTAGCTTGAAAA-----AAAAAAAAACCCACCCTAAAG-TGG--GAACTT--------
VoleM.a. AGTAGCTTGAAAA-----AAAAAA----CCCACCCTAAAA-TGG--GAACTT--------
Mouse AGTAGCTTGAAAATTCCTAAAGTGGGAATTTATTTTAGAGATGATAGAACCTGTTTCCCC
Coastmole AGTATTCCGAA----------GCA-----CCACACTCAAC-TGA--GGATTTAG------
Human AGTAGTTTGAA----------G
VoleM.r. AAATAATAT------ATATATATATCCAGATCTAATCATTC--TTGAGAGTACACTTCAA
VoleM.t. AAATAATAT------ATATATATATCCAGATCTAATCATTC--TTGAGAGTACACTTCAA
VoleM.a. AAATAATAT------ATATATA--TCCAGATCTAATCATTC--TTGAGAGTACACTTCAA
Mouse ACTTTACATTTTAAAATATGTCTGCCAGGATCTAATCATTCCTTTAAACGTACACTTCAA
Coastmole AACTGATG-------------------GGA-CCAGTTTTTC---------TTTGTTATAT
Human AAATGATA-------------------AGA-CGAGTTCC-C---------TATTTTATAA
VoleM.r. AGAGAGATTCGCCTATTAAGAAAATATCT----GTCTCTATTGGCGCCGAGTATTTA--A
VoleM.t. AGAGAGATCCGCCTATTAAGAAAATATCT----GTCTCTATTGGCGCCGAGTATTTA--A
VoleM.a. AGAGAGATCCGCCTATTAAGAAAATATCT----GTCTCTATTGGCGCCGAGTATTTA--A
Mouse AGAGAGATTTTCCTAGTAAGAAAAGAGCT----TTCTCTAGTG-TGAAGGGTGCTTTGTA
Coastmole TAAAAGA----------AAG-----ATC------TTTTTG--------GGGTAT------
Human GAAAAAT----------AAGCCAAAATTAAATATTCTTTT--------GGATAT------
VoleM.r. GCTTTCTGGT-TTTCGTTCTTT------AATTATTGTGCATAAATGAGTGTGTA------
VoleM.t. GCTTTCTGGT-TTTCGTTCTTT------AATTATTGTGCATAAATGAGTGTG--------
VoleM.a. GCTTTCTGGT-TTTCGTTCTTT------AATTATTGTGCATAAATGAGTGTGTGTGTGTG
Mouse GCCGCCGAGTACTTAGGTCTTTTTTGGGAGCTATTGTGTA----TGAGTGTA--------
Coastmole ----------AATTTTTA-GCA---GTGAGT-T---------
Human ----------AAATTTCA-ACA---GTGAGA-TA--------
CMXISTREV5
VoleM.r. ----TGTGTGTGTGTGTGTGTGCATGTTTGCATGTGCACTTGAGTATCTGTACTCGTG
VoleM.t. ----TGTGTGTGTGTGTGTGTGCATGTTTGCATGTGCACTTGAGTATCTGTACTCGTG
VoleM.a. TGTGTGTGTGTGTGTGTGTGTGCATGTTTGCATGTGCACTTGAGTATCTGTACTCGTG
Mouse ----TGTATGTGTGTGTGTACATGCATGTT-GC-TGCGCGC----------AGTCATT
Coastmole -----GT-TGCCTG-ATAAAAGTGAAT-----CATATTC----------CAGCCTCT
Human -----GC-TGCCTA-GTGGAAATGAAT-----AATATCC----------CAGCCACT
VoleM.r. CATTTGCATTGGTGACCAGAGAGCAATGGGAACTGGTTC---TACCTTGTGGGTCCCTGG
VoleM.t. CATTTGCATTGGTGACCAGAGAGCAATGGGAACTGGTTC---TACCTTGTGGGTCCCTGG
VoleM.a. CATTTGCATTGGTGACCAGAGAGCAATGGGAACTGGTTC---TACCTTGTGGGTCCCTGG
Mouse CATTCACAT-GGTGCTCAGACAACAATGGGAGCTGGTTCGTCTATCTTGTGGGTCC-TGG
Coastmole TGTATACAG-GCTGTTTGG-TAGCA--GAGAAA----------ATTTAGTAAG-------
Human AGTGTACAG-GGTGTTTTG-TGGCA--CAGGATT---------ATGTAATATG-------
VoleM.r. AGATAAAAACTCAGATCGTCAGGTTTGGCAGCAAGTGCCTTCACTCTCTGAG--CCATCT
VoleM.t. AGATAAAAACTCAGATCGTCAGGTTTGGCAGCAAGTGCCTTCACTCTCTGAG--CCATCT
VoleM.a. AGATAAAAACTCAGATCGTCAGGTTTGGCAGCAAGTGCCTTCACTCTCTGAG--CCATCT
Mouse AGATCAAAG-TGAGATCATCAGGCTTGGCAGCAAGTGCCTTTACCCTCCGCGTGCCATCT
Coastmole ------GAACT----------GCTGA--AACAAGTACCT----------AG--TTATC-
Human ------GAACT----------GCTCA--AGCAAATAACT----------AG--TCATC-

```
VoleM.r.        TGCCAGCCC---------------ATGTGACATTGCTGATGAAACCA-------------
VoleM.t.        TGCCAGCCC---------------ATGTGACATTGCTGATGAAACCA-------------
VoleM.a.        TGCCAGCCC---------------ATGTGACATTGCTGATGAAACCA-------------
Mouse           TGCCATCCCGCTGCTGAGTGTTTGATATGACATTGCTGATGAAAATAATCATCACAACAG
Coastmole       ----------------------------------------A-----------------G
Human           ----------------------------------------ACAACA------------G
                                                         * *  *

                 CMXISTREV6➔
VoleM.r.        CAGTTCTTC-AA-GTTTCTGAAAAAAAAAAAAAAACAATATTTTTCTGAGAAGGATGT-CA
VoleM.t.        CAGTTCTTC-AAAGTTTCTGGAAAAAAAAAAAA--AATATTTTTCTGAGAAGGATGT-CA
VoleM.a.        CAGTTCTTC-AAAGTTTCTGAAAAAAAAAAAAC--AATATTTTTCTGAGAAGGATGT-CA
Mouse           CAGTTCTCCCAGCATTACTGAGAAATGA------TACTATTTTTCTGAGGAGGATGTTCA
Coastmole       CAGTTCTTT-GTAATCNCTGAAAAAGGA------TACTATTCCTCTNAGAAGGATGC-CN
Human           CAGTTCTTT-GTAATAACTGAAAAAGAA------TATTGTTTCTCGGAGAAGGATGT-CA
                *******       *  ***  ***  *       * * **  **  ** ******  *

VoleM.r.        AGAAA-TCGACCCAGTT
VoleM.t.        AGAAA-TCGACCCAGTT
VoleM.a.        AGAAA-TCGACCCAGTT
Mouse           AGTAACTCATCCCAGTG
Coastmole       AAANA-TCCGCCCAGCT
Human           AAAGA-TCGGCCCAGCT
                *   * **  *****
```

	Coast mole	Human	Mouse	Vole M.r.	Vole M.a.
Human	77				
Mouse	44	54			
Vole M.r.	40	51	72		
Vole M.a.	39	51	72	98	
Vole M.t.	40	51	73	98	98

Figure 13a: Aligned sequences at the end of exon one.
The alignment of Coast mole *Xist* sequence at the end of exon one with the corresponding *Xist* fragments from human (M97168; bases 10011-11295), mouse (L04961; 7861-9422), vole *Microtus rossiameridionalis* (AJ310130; 6325-7889), vole *M. arvalis* (AJ310129; 6516-8080) and vole *M. transcaspicus* (AJ310127; 6331-7893). Primers derived from sequence conserved between the human gene and a short bovine EST (BE483406; not shown) are green, while those that are mole-specific are red. Note the flanking *AluI* sites highlighted in turquoise.

```
Coastmole       AGCTGCACTTATCTTCATCCTGCCCATCAGACATGGACCCTTCCACGTCCTTCTTTGCAT
Human           AGTTGCACTTTCCTTGGTCC-ACCCATTATACATGAACCCCTCTACTTCCTT--TCGCAT
                ** *******  ***  ***  ***** * ***** **** ** ** *****  * ****

Coastmole       TGCTTCTGAGCAGTGCTGACTACCCAAAGCCCC--CTGTGTTCTTAATTAACACAGTAAT
Human           TGCTTCTGAGTA-TGCTGACTACCCAAAGCCCCTTCTGTGTTATTAATAAACACAGTACT
                ********** * ********************  ******* ***** ********* *

Coastmole       G----TCCCATTTTTCAGCCCATCGGCCCAAGATCTCCATGT--------TGTGTATGTG
Human           GATTGTCCCATTTTTCAGCCCATCAGTCCAAGATCTCCCTACCACTTTGGTGTGTTGGTG
                *    ****************** * ********** *          *****  ***

Coastmole       CGGTGTTGACTACCAAAAGCAGGCCGAAACTCGGTGGATGGACCCTCCTTCCTTTTCCTG
Human           CAGTGTTGACTATGAAAAGCAGGCCTGAACTAGGTGGATAAGCCTTCACTCATTTTCTTT
                * **********  ***********  **** *******   ** **  ** ***** *

Coastmole       CATTTGTTAATGATTCCAACACCAGTC---GTCAAATTCTGGGGGCAGGAACCATCCTTG
Human           CATTTATTAATGATCCTAGTTTCAATTATTGTCAGATTCTGGGGACAAGAACCATTCTTG
                ***** ******** * *    ** *    **** ********* ** ******* ****

Coastmole       CCCACCTCTGTTACTGCTTTACTGGGCAAAATGCCCAAGATAAGCCAGGCCCACAGAACT
Human           CCCACCTGTGTTACTGCTTTACTGTGCAAAATACTGAAGGCAAGTCAGACCCAGGGAGCT
                ******* ********************* ******* *  ***  *** *** ****  ** **

Coastmole       GGATTGCTGTCTTTTATTTTGGATGTTCAGTAGTTACCATTATCAGTTGAAAATTGTCTC
Human           GGATTGCCATCCTTTATTTTGTGTTTCCAGTG--TAC-ACTAT------AAAATTGTCTC
                *******  ** *********  * * ****   *** * ***      ***********

Coastmole       CCCAGGAAAGAAG-TTAGCGTTGTCTCT----------TTCCTGAGCAGAGTGCCTGGCT
Human           CCCAGGAAGGAAGGTTGGCACTTTCTCTGCATTCTTCTTTCCAGAGCAGATTGCCTGGTT
                ******** **** ** **  * *****          **** ******* ******* *

Coastmole       GTAAATCTCTCATC-------TTATATATTGCTATTGTAAAGTGCCAGTTGCCAGAG-GC
Human           AAGAATCTCTTGTTGTCCCTTCTGTATATTGTTATTGTAAAGTGCCAAATGCCAGGATAC
                   *******  *         * ******* ***************  ******    *

Coastmole       AACCA--AAAGTTGCTATTAATTTTTTA-----TTTTCTAAAGAAAGGTATCTGGAATAT
Human           AGCCAGAAAAATTGCTTATTATTATTAAAAAAATTTTTTTAAGAAAGACATCTGGATTGT
                * ***  *** *****  * *** ** *     **** * *******  ******* * *

Coastmole       CAGGTGGAGTTGATAACCTGGTCATTAAATATTTGAAGTCCTAAAACCCATTTATATCAT
Human           AGGGTGGACTCGATAACCTGGTCATTATTTTTTTGAAGCCAAAATATCCATTTATACTAT
                  ****** * ****************  * ******* *  ** * *********  **

Coastmole       GTACCTAATGACCAGTGTCTCTCATTTTA-CTAAGGGTGGTGGGTCTGNGGATAGACCAC
Human           GTACCTGGTGACCAGTGTCTCTCATTTTAACTGAGGGTGGTGGGTCTGTGGATAGAACAC
                ******  ********************** ** *************** ******* ***

Coastmole       TGTGACTCTTGCTACTTTATTATTGCTCAAGTTAGTGAGTTTTCAAGTGAAACTCTTAAA
Human           TG--ACTCTTGCTATTTTAATATCA---AAGATA-----TTCTAGAGTGGAACTCTTAA-
                **  ********** **** ***     *** **     ** *  **** *********

Coastmole       GACCGGTACCTTTG----GGCTCTACCACCATTCATTTTTAGAAAAACTATGGAAACTTT
Human           GACCAGTATCTTTGTGTGGGCTTTACCAGCATTCACTTTTAGAAAAACTACCTAAATTTT
                **** *** *****    **** ***** ****** **************   *** ***

Coastmole       ATAGTCCTTAAATTTCTTCATCTGGAGCACCAACCAGTCCAGTCCTTATTTATTTCAATA
Human           ATAATCCTTTAATTTCTTCATCTGGAGCACCTGCC---------CCTACTTATTTCAAGA
                *** ***** **********************  **         * ** ********* *

Coastmole       AGATTGTTATAAAACAATTATATGAGAGAATGAACATATGCTGAGGTGCTTTTGAAAACC
Human           AGATTGCAGTAAAACGATTAAATGAGG----GAACATATGCAGAGGTGCTTTTAAAAAGC
                ******   ****** **** *****     ********** ********** **** *

Coastmole       ATAGATCACCTTCTATGATTAATCTATTACATAAAATGCACCCTGTTTTAATTATGCTAA
Human           ATATGCCACCTTTTTT-ATTAAT-TATTATATAAAATGAAGC---ATTTAATTATAGTAA
                ***   ****** * * ****** ***** ******** * *    *********  ***
```

```
Coastmole       T--TNTAAAGTATTCCGAAGCACCACACTCAACTGAGGATTTAGAACTGATGGGACCAGT
Human           TAATTTGAAGTAGTTTGAAGTACCACACTGAGGTGAGGACTTAAAAATGATAAGACGAGT
                *  * * ***** *  **** ******** *  ****** *** ** ****  *** ***

Coastmole       TTTTCTTTGTTATATTAAAAGAAAG-----ATC------TTTTTGGGGTATAATTTTTAG
Human           TCC-CTATTTTATAAGAAAAATAAGCCAAAATTAAATATTCTTTTGGATATAAATTTCAA
                *   ** * *****  ****  ***     **       * *** ** ***** *** *

Coastmole       CAGTGAGTT-GTTGCCTGATAAAAGTGAATCATATTCCAGCCTCTTGTATACAGGCTGTT
Human           CAGTGAGATAGCTGCCTAGTGGAAATGAATAATATCCCAGCCACTAGTGTACAGGGTGTT
                ******* * * *****  *  ** ***** **** ****** ** ** ****** ****

Coastmole       TGGTAGCAGAGAAA-ATTTAGTAAGGAACTGCTGAAACAAGTACCTAGTTATCAGAACAG
Human           TTGTGGCACAGGATTATGTAATATGGAACTGCTCAAGCAAATAACTAGTCATCACAACAG
                * ** *** ** *  ** ** ** ********* ** *** ** ***** **** *****

Coastmole       CAGTTCTTTGTAATCNCTGAAAAAGGATACTATTCCTCTNAGAAGGATGCCNAAANATCC
Human           CAGTTCTTTGTAATAACTGAAAAAGAATATTGTTTCTCGGAGAAGGATGTCAAAAGATCG
                **************  ********* *** * ** ***  ********* * *** ***

Coastmole       GCCCAGCT
Human           GCCCAGCT
                ********
```

Figure 13b: Aligned sequences at the end of exon one, excluding the rodents.
An alignment of the Coast mole Xist fragment located at the 3' end of exon one with the corresponding human sequence. The rodent genes were removed to highlight the conservation between the Coast mole and human genes. For primer locations and levels of sequence similarity see Part A of this figure.

```
Coastmole       TTTTCTTTGTTATATTAAAAGAAAG-----ATC------TTTTTGGGGTATAATTTTTAG
Human           TCC-CTATTTTATAAGAAAAATAAGCCAAAATTAAATATTCTTTTGGATATAAATTTCAA
                *   ** * *****  ****  ***     **       * *** ** ***** *** *

Coastmole       CAGTGAGTT-GTTGCCTGATAAAAGTGAATCATATTCCAGCCTCTTGTATACAGGCTGTT
Human           CAGTGAGATAGCTGCCTAGTGGAAATGAATAATATCCCAGCCACTAGTGTACAGGGTGTT
                ******* * * *****  *  ** ***** **** ****** ** ** ****** ****

Coastmole       TGGTAGCAGAGAAA-ATTTAGTAAGGAACTGCTGAAACAAGTACCTAGTTATCAGAACAG
Human           TTGTGGCACAGGATTATGTAATATGGAACTGCTCAAGCAAATAACTAGTCATCACAACAG
                * ** *** ** *  ** ** ** ********* ** *** ** ***** **** *****

Coastmole       CAGTTCTTTGTAATCNCTGAAAAAGGATACTATTCCTCTNAGAAGGATGCCNAAANATCC
Human           CAGTTCTTTGTAATAACTGAAAAAGAATATTGTTTCTCGGAGAAGGATGTCAAAAGATCG
                **************  ********* *** * ** ***  ********* * *** ***

Coastmole       GCCCAGCT
Human           GCCCAGCT
                ********
```

Figure 13b: Aligned sequences at the end of exon one, excluding the rodents.
An alignment of the Coast mole Xist fragment located at the 3' end of exon one with the corresponding human sequence. The rodent genes were removed to highlight the conservation between the Coast mole and human genes. For primer locations and levels of sequence similarity see Part A of this figure.

Based on the organization of the human *XIST* gene, the distance between the two mole fragments should be approximately 8 kilobases (kb), a size that is within the range of the ExpandLong PCR system. PCR using this kit and several Coast mole primer combinations succeeded in amplifying a band in the 12 kilobase range (Figure 11, Figure 14) in genomic DNA (54° C annealing, 8 minutes extension, 40 cycles). I tried to directly sequence this band, but the amount of template necessary to sequence a product of this size proved impossible to obtain through PCR and sequencing will require cloning of the product. Once the known sequence separating the primers (1.5 kb) was accounted for, the length between the end of the two fragments was reduced from over 12 kilobases to approximately 10.5 kilobases. Thus, barring the presence of an intron, the length of mole exon is at least 13 kilobases. Keeping in mind that neither the start of exon one, nor the 5' repeat region identified by Hendrich et al. (1993), have been reached and that these regions may extend another kilobase or two, the first exon of mole *Xist* may extend over 15 kilobases. Since this exon, which forms the majority of the *Xist* RNA, in mouse, human and voles ranges between 8 and 11 kilobases, the theoretical size of the mole equivalent is very unusual. Although there is a possibility that the 12 kb fragment is a PCR artifact, it is doubtful since the reaction has been replicated a number of times, with different primer pairs in the region and under relatively stringent conditions (54° C annealing). Another explanation for this increased size may be due the presence of an intron in this region of the mole gene, which is not present in other species. To address this problem SuperScript reverse transcription (RT) reactions were performed on Coast mole RNA of good quality (CM16, CM17) in the hope that very large cDNA molecules may be produced, although the upper limit of the enzyme is only 12.3 kb. If an intron were present, a change in size of the PCR fragment would be noticeable depending on whether DNA or cDNA was used as a

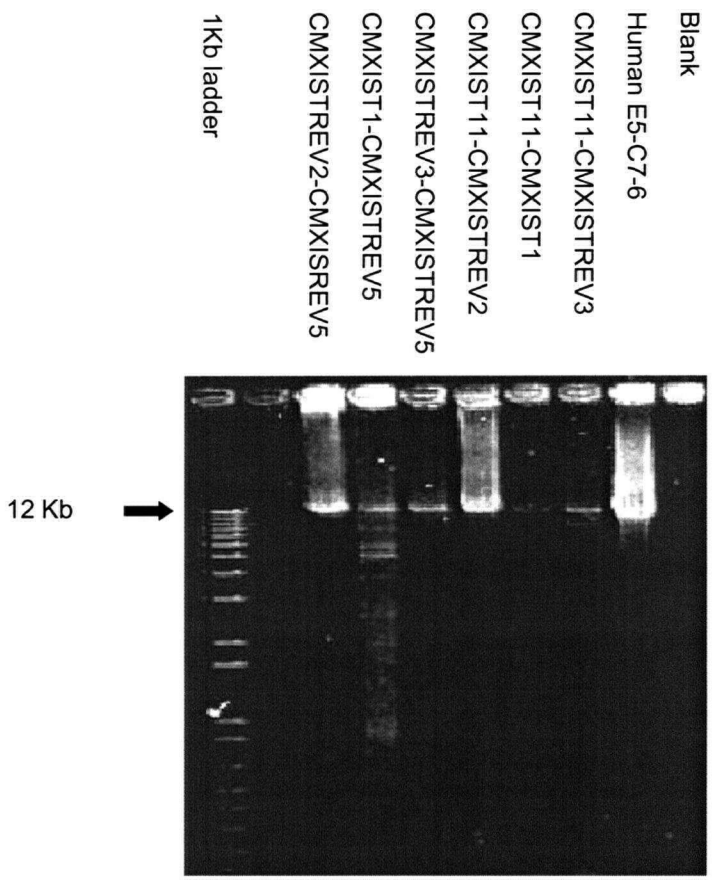


Figure 14: The 12kb Coast mole fragment produced by ExpandLong PCR.
The PCR was performed with various primers from the two fragments of Coast mole *Xist* using an annealing temperature of 54C and 40 cycles of amplification.The sequences and locations of the Coast mole primers are presented in Table 2 and Figures 12 and 13. Human DNA (GM0773) in combination with the E5-C7-6 primer pair (Table 3) was used as a positive control.

template. However, ExpandLong PCR on SuperScript RT cDNA template using the CMXIST5REVERSE and CMXIST2REVERSE primers, which gave the strongest 12 kb fragment (see Figure 14), did not result in a product. The possible interpretations of this result will be addressed in the Discussion of this chapter.

Another approach that was considered in the attempt to determine the approximate size of the Coast mole *Xist* gene, was northern analysis with a newly generated mole *XIST* probe. RNA of high quality was available from the female cell lines, but the male cell line grew very poorly and could not provide the amount of RNA necessary for this method. Since the male RNA was essential as a negative control, I did not proceed with the northern blot.

Discussion

PCR with primers derived from conserved regions of the human, bovine, equine and lepine *Xist* genes was successful in detecting the Coast mole homologue. This PCR method was quite efficient in amplifying mole sequences, especially since *Xist* is an untranslated RNA and does not exhibit the high levels of conservation seen in protein-coding genes (Hendrich et al., 1993). However, while conservation between cow, horse, human and rabbit was a good indicator of identity in moles, there were primer pairs that, despite being perfectly conserved between these species, failed in mole. When the sequence around one of these pairs (SKXIST1,2) was obtained by inverse PCR, it was shown that the mole gene showed several mismatches in the region explaining the PCR failure in this case. Using conserved primers, along with inverse PCR over 2.5 kilobases of the gene were identified and sequenced. The alignment of all known genes and gene fragments are presented in Figure 15. The Coast mole *Xist* fragments show significant similarity (around 70-80%) to their counterparts in all the examined species, with the exception of the rodents, represented by the mouse and three vole species, in which the sequence identity falls to 20-50%. The initial sequences obtained are biased towards conserved regions of the gene, since they were derived from conserved primers. However, the inverse PCR fragments should not be affected by this bias, since they do not rely on conserved primers. The strikingly low identity with the rodent sequences agrees with the work of Hendrich et al. (1993), where mouse showed the least sequence conservation compared to the human, bovine, equine and lepine genes. The similarity scores seen for the Coast mole fragment fall within the range seen for the other sequenced species in my alignments and those reported in the literature (Brockdorff et al., 1992, Brown et al., 1992, Hendrich et al.,1993). Thus, the new mole

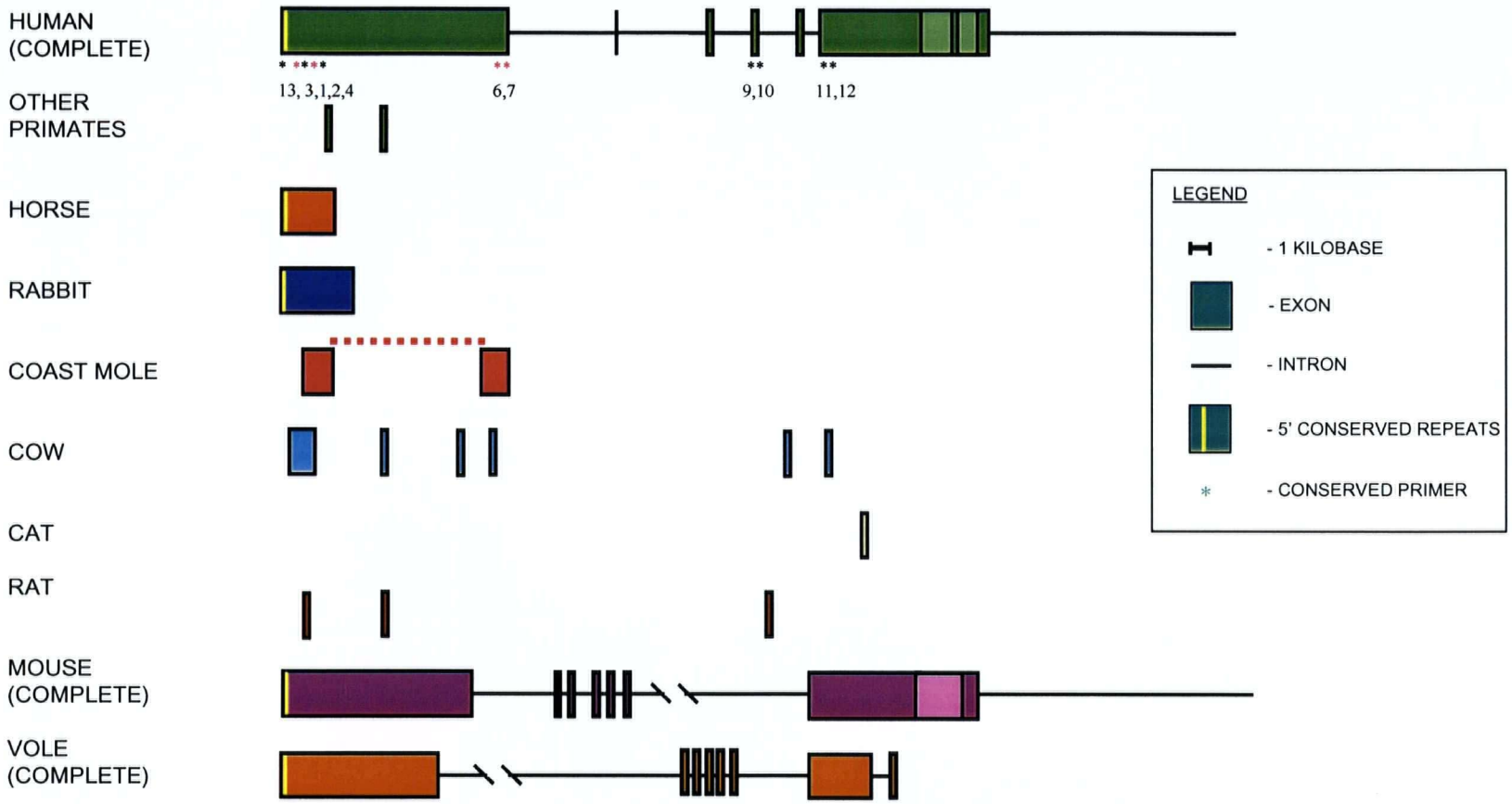


Figure 15: The alignment of known *Xist* sequences as of April 10th 2001 including the Coast mole fragments. The Coast mole *Xist* sequences are shown in red and correspond to the very 5' and 3' regions of human exon one. The dashed line above the Coast mole fragments represents the 12 kb PCR fragment (see text). Conserved *Xist* primers are shown as asterisks as in Figure 8.

sequence is similar enough to be an *Xist* homologue, yet shows significant differences and unique base changes to prove that its presence was not due to contamination, by human DNA for example.

Expression of both of the fragments of Coast mole *Xist* was only detected in female tissue and cell lines and not in the male. This mode of expression is typical of *Xist*, as the gene is only expressed from the inactive X chromosome, and confirms that the Coast mole fragment is an expressed, and probably functional homologue of this gene. *Xist* is expressed in the human and murine testis, but the situation in moles could not be established since RNA from this tissue was not available.

While the complete *Xist* gene in *Scapanus* has not yet been sequenced, it seems that it will be of similar, if not greater, size than the mouse, human and vole homologues. The 12 kilobase fragment obtained through long PCR indicates that, in the absence of an intron, Coast mole exon one may extend over 15 kilobases, which is 4 kilobases greater than its 11 kilobase human counterpart (Brown et al.,1992). Exon one in mouse and voles is even smaller at approximately 8 kilobases (Brockdorff et al., 1992, Nesterova et al., 2001 *in press*). This unexpected size of the mole exon could be due to an intron in this species. Unfortunately, the long PCR on SuperScript cDNAs, which would have provided an answer to this question, did not work. The failure of this PCR can be due to one of two reasons: if the exon is indeed 12 kilobases, it is not surprising that a cDNA of that size was not produced, since the empirical upper limit of the SuperScript reverse transcriptase is 12.3 kilobases. Even if a small intron were present, reducing the cDNA size to 8 to10 kilobases, it is still doubtful that the reverse transcriptase would have been able to copy a fragment of that length since we have only been able to produce

a five kilobase transcript in our laboratory (J. Chow, personal communication). The combination of these uncertainties renders the absence of a cDNA band in this experiment inconclusive.

The Coast mole *Xist* fragments do not contain any significant protein-coding potential, although a number of small open reading frames were scattered throughout the sequence, as was the case in the murine and human genes (Brockdorff et al., 1992, Brown et al., 1992, Hendrich et al., 1993). The three Coast mole ORFs with excellent coding potential in the 5' region of exon one all overlapped with each other and with human ORF4 (Brown et al., 1992, Hendrich et al, 1993) in different reading frames, which would not be expected if there was a conserved polypeptide in the region. Based on the negative results of the exhaustive searches in human and mouse *Xist*, which failed to identify a functional ORF, I did not perform in depth analyses of the ORFs in the mole *Xist*.

The pattern of *Xist* sequence conservation is unlike that seen for most protein coding genes in which large domains are well conserved. Instead, small blocks of identity are surrounded by sequence that has been poorly conserved (Brockdorff et al., 1992, Brown et al., 1992, Nesterova et al., 2001 *in press*). The functional significance of these conserved areas, such as the 5' repeats has not been established, since the *Xist* knockouts to date have not been targeted to small regions, but have involved rather large portions of the gene (Penny et al., 1996, Marahrens et al., 1997). Since neither the putative region of the 5' conserved repeat region, nor any of the other major repeat areas (Brockdorff et al., 1992, Brown et al., 1992), have been sequenced in *Scapanus*, it is not known if these motifs are present. However, even in the small fragments that have been identified, the unique pattern of *Xist* conservation can be seen. In both fragments, we

see small regions of conservation, some of which seem to be imperfect repeats, interspersed with divergent sequence.

Xist is almost certainly a functional RNA (Brockdorff et al., 1992, Brown et al., 1992) as it contains no significant open reading frame and does not leave the nucleus - a requirement for translation. Although the function and secondary structure of the transfer RNAs (tRNAs) and ribosomal RNAs have been extensively studied, the presence and significance of other non-coding RNAs has become apparent only recently and not much is known about how they perform their tasks. RNAs are now known to be involved in the regulation of gene expression by mediating mammalian X-inactivation, dosage compensation in *Drosophila* and the developmental regulation of certain genes in *C. elegans*. Functional RNAs are also involved in processing and modification of tRNA, rRNA and mRNA and include the RNAs involved in splicing, RNA editing and tRNA maturation. Additionally, there are the "housekeeping" RNAs, such as the telomerase RNA and the signal recognition particle RNA (SRP) involved in protein sorting in the cell (reviewed in Caprara and Nilsen, 2000). Although RNA folding is still poorly understood, it is now clear that functional RNAs adopt a remarkably diverse set of secondary and even tertiary structure interactions that surpass the stem-loop interactions envisioned previously (reviewed in Caprara and Nilsen, 2000). While the mode of action of some functional RNAs is becoming clearer, the mechanism through which the *Xist* RNA mediates X-chromosome inactivation is still a complete mystery, although the LINE interaction hypothesis is a good candidate (see Chapter 1).

The elucidation of *Xist* function is hampered by the fact that, at least at the primary sequence level, the RNA doesn't show homology to almost any sequences in the database. One interesting exception are the *Xlsirt* RNAs (*Xenopus laevis* short interspersed repeat transcripts)

in the *Xenopus* species, which are translocated with the germ plasm in the oocyte and may contribute to the establishment of the germ cell lineage (Kloc et al., 1993). The region of homology between *Xist* and the *Xlsirt* RNAs is very small (26 nucleotides) and consists of a repeated sequence in the *Xenopus* gene that is similar to the core of the 5' conserved repeats in human, mouse, rabbit, horse and voles. The occurrence of these sequence motifs in two classes of possibly functional RNAs is probably not a coincidence, but at this time the function of the repeats remains unclear.

The view that *XIST/Xist* functions at the RNA level is strengthened by recent analyses of the non-coding *H19* RNA and the telomerase RNA. The function of the *H19* RNA is still an enigma, although it is known to be involved in the negative regulation of the imprinted *Igf2* gene through the competition for a shared enhancer (reviewed in Juan et al., 2000). Additionally, induced expression of *H19* leads to the suppression of cellular proliferation and tumourigenicity in certain malignant cell lines, prompting the idea of a new role as a tumour suppressor (reviewed in Juan et al., 2000). Since no substantial open reading frames have been identified in *H19*, it was initially thought that the RNA was a non-functional by-product of enhancer competition with *Igf2*. However, a study by Hurst and Smith (1999) examining the sequence conservation of *H19* between rats and mice showed that the RNA is subject to stabilizing selection and probably functional. A recent analysis of the secondary structure of *H19* RNA in a number of mammals (cat, lynx, gopher, elephant, orangutan, cat, mouse, rat, rabbit and human), observed evolutionary conservation of the secondary structure, despite low sequence identity (Juan et al., 2000). A virtually identical conclusion was reached by Chen et al. (2000) in their examination of telomerase RNA, which serves as the RNA template for the telomerase ribonucleoprotein enzyme that maintains telomere length in organisms with linear

chromosomes. The authors compared the telomerase RNA from the five classes of vertebrates (mammals, reptiles, birds, amphibians and fish) and found that all the RNA molecules folded into similar secondary structures even though primary sequence conservation was negligible (for example human and shark only showed 44 % nucleotide identity). The structure of the vertebrate telomerase RNA also showed a remarkable similarity to the architecture of the ciliate telomerase RNA, although some domains are different. The secondary structure of the *XIST/Xist* molecule has not been established, primarily due to the large size of the molecule and insufficient computing power. However, it is interesting to note that the pattern of primary sequence conservation seen for *H19* and the telomerase RNAs, in which blocks of high identity are surrounded by areas of low conservation, is very similar to that seen for *Xist*. (this study, Brockdorff et al., 1992, Brown et al., 1992, Hendrich et al., 1993, Nesterova et al., 2001 *in press*). Thus the mode of *Xist* sequence conservation may simply reflect the fact that it is the secondary structure, rather than the primary nucleotide sequence, that is the subject of natural selection. If the presence of strategically placed repeats or blocks of conservation is sufficient to ensure the proper secondary structure, the intervening sequence need not be conserved. However, since all known *Xist* genes seem to be quite large, it may be that the approximate size of the RNA is also under selective pressure. It is possible that a long RNA molecule is necessary to be able to effectively bind the hypothetical anchoring repeat sequences and then coat an entire chromosome.

Although the *Xist* gene does not seem particularly conserved amongst the eutherians, the extremely low homology exhibited by the rodent sequences with respect to the other species in the regions studied is striking. While the human, mole, rabbit, horse and cow fragments show 70-80 % identity, the mouse and vole scores are as low as 20 % for the 5' region of exon one

and are slightly better at 40-50 % for the end of exon one. Even amongst themselves these muroid rodents do not show a lot of conservation in the tested regions; although the vole genes are 98 % identical, the mouse has only 70% similarity to the voles. While rodents have some blocks of conservation shared by other species (see Figures 9, 12, 13) and share the conserved 5' repeats with humans, cows, horses and rabbits (Brockdorff et al., 1992, Brown et al., 1992, Hendrich et al., 1993, Nesterova et al., 2001 *in press*), the surrounding sequence shows a high degree of divergence. The functional significance, if any, of this lack of conservation is unknown since *Xist* function may only depend on a small proportion of the entire sequence, such as the conserved 5' repeats or other conserved blocks. For example, a YAC containing human *XIST* can induce inactivation in mouse autosomes, although without the accompanying late replication and histone hypoacetylation, indicating that, despite the general lack of conservation, crucial motifs may have been preserved (Heard et al., 1999, Migeon et al., 1999).

A possible source of the sequence divergence seen in the mouse and voles may be the higher rate of nucleotide substitutions reported in rodents in comparison to primates (Wen-Hsiung et al., 1996). According to this hypothesis, which still remains controversial, organisms with a short generation time should go through more cycles of DNA replication per unit time and should have a higher mutation rate than organisms with longer generation times. Insectivores, especially the shrews, may be superficially similar to the muroid rodents, but they do not share their prolific breeding habits and thus have longer generation times (Gorman, 1990, Nagorsen, 1996). Although small body size has been correlated with an increased nucleotide substitution rate, due to a higher metabolic rate and generation time, a phylogenetic study in shrews based on complete mitochondrial DNA does not show an increased rate in nucleotide substitution (Fumagalli et al.,1999). Since all of the organisms examined in our study of *Xist*

have longer generation times than rodents, the extreme divergence seen in the mouse and vole is consistent with an increase in nucleotide substitution in this lineage. Of course, it must be kept in mind that I have analyzed a very small region (2.5 kb) out of a very large RNA and that the pattern of conservation may differ in other regions of the gene. Before any firm conclusions regarding the levels of conservation and can be made, it is necessary to completely sequence *Xist* in several organisms, including the mole. Additionally, more precise deletions in the gene will be necessary to identify functionally significant regions.

The data presented in this chapter are the first to show the presence and female-specific expression of the *Xist* gene in an insectivore. Although functional studies have not been performed in *Scapanus*, the female-specific expression pattern in combination with the relatively high sequence conservation are good indicators that this gene is involved in X-chromosome inactivation in the Coast mole.

Chapter 5

Discussion and future directions

The methylation analysis of Coast mole genes, along with results of Jegalian and Page (1998), demonstrate that both X-chromosome inactivation and silencing of X-linked sequences through methylation have been conserved in the Insectivora. While the *ZFX, ARA* and *FMR1* genes were presumed to be on the X chromosome of the mole on the basis of Ohno's law (1967), the consistent methylation differences between the male and female alleles, a situation only seen for X-inactivated genes, confirm this prediction. The results also highlight the extraordinary conservation of the CpG island regulatory regions located in the 5' regions of *ZFX, ARA* and *FMR1* and suggest the importance of strict dosage compensation for the studied genes. The status of the *ZFX* gene, which escapes inactivation in the Coast mole and European hedgehog (Jegalian and Page, 1998), supports the observation that this gene became silenced only in the rodent lineage (Jegalian and Page, 1998). It would be interesting to establish whether *ZFY* is present and functional in the mole and thus strengthen the hypothesis that it is the loss of the Y homologue that led to the inactivation of *ZFX* in the muroid rodents.

A full appreciation of the nucleotide conservation of the mole *Xist* gene will be gained once the entire gene is sequenced. At the moment I am attempting to clone the 12 kilobase fragment (Chapter 3) into a plasmid vector. Due to the rather large size of this product, it may be necessary to fragment it by performing restriction digests and then clone the smaller pieces. A parallel strategy is the application of inverse PCR (Fig 3), which may allow the sequencing of the 5' end of the gene and the 12 kilobase central fragment. This strategy will probably not be

successful in isolating the 3' terminus of the gene, since it is reasonable to expect that the coding sequence will be interrupted by several introns. In fact, the Coast mole fragment at the 3' region of exon one seems to be at very end of that exon. Therefore, the 3' region of the gene will probably have to be amplified by 3' RACE (rapid amplification of cDNA ends) using polyT and internal gene primers. The polyT primer anneals to the polyA tail of polyadenylated RNAs, such as *XIST/Xist* (Brockdorff et al., 1992, Brown et al., 1992), and in combination with a primer within the known gene sequence, the intervening RNA is amplified and the lengthy introns are avoided. I have attempted this technique once, without any conclusive results (data not shown). Additionally, EST sequencing of various species may allow the design of more conserved primers in the future. In the case that all of these PCR-based methods fail, it will be necessary to construct a mole cDNA library if Coast mole *Xist* is to be sequenced.

Preliminary analyses of the Coast mole *Xist* fragments, show that the gene is quite similar to the available sequences, with the exception of the rodents, although it will be necessary to analyze the entire sequence in order to establish global homology levels. An encouraging result of the analyses presented in this work is the detection of conserved regions that are not apparent when only the human and murine genes are compared. Confirmation that these are functional domains, along with the previously identified 5' conserved repeats, would require deletions of these regions.

It is unclear whether Coast mole *Xist* will aid in the search for a marsupial homologue at this point, precisely because of the relatively high levels of sequence similarity seen between most of the genes. However, the complete sequence of the *Scapanus* gene will certainly be helpful in determining the presumably conserved secondary structure of the *XIST/Xist* RNA. The elucidation of this structure could clarify the mechanism of *XIST/Xist* function, which may

involve the formation of a ribonucleoprotein, possibly analogous to the MSL complex in *Drosophila*, that establishes the closed chromatin conformation of the mammalian inactive X (Kelley and Kuroda, 2000a,b). Despite the continuing ignorance regarding *Xist* function, it is becoming increasingly clear in the case of non-coding RNAs that the secondary structure is the phenotype acted on by natural selection. Studies of the *H19* RNA in several eutherian mammals has identified a secondary structure conserved in all the species examined (Juan et al., 2000). Even more striking are the results of Chen and colleagues (2000) who identified a conserved secondary structure of the telomerase RNA structure that was conserved not only within mammals and vertebrate, but down to the protozoan ciliates. Perhaps searches of the marsupial genomes, which have yet to be sequenced, for a conserved secondary, rather that primary, structure of *Xist* will yield a functional homologue of the eutherian gene.

While I believe that it is necessary to expand our studies of X-inactivation beyond the human and mouse models to fully understand the mechanism and evolution of this process, it is doubtful that most of the studies will be done in the mole. As a result of the mostly underground existence of the majority of moles, little is known about the breeding habits and genetics of these animals. Moles have not been bred in captivity and no embryonic stem cell lines, which would be necessary for performing *Xist* knockout experiments, are available. Although it is possible to keep a mole in captivity, where they get used to handling and make "gentle, but demanding" pets, breeding programs are almost certainly impossible (Gorman and Stone, 1990). Moles are solitary territorial animals and two captured individuals placed in the same container have been known to fight to the death (Gorman and Stone, 1990). Only during the short mating season can the moles tolerate each other's presence for limited periods of time. Furthermore, it is unlikely the genetics of this species will be intensively studied in the near future since most of

the attention moles get from humans seems to be negative. The common moles are usually viewed as pests and since they have little or no commercial value, they are eliminated in various inhumane ways, including traps, poison and "mole-vacuums", which suck the animals out of their tunnels. A temporary, and rather unfortunate, increased interest in these little insectivores occurred at the beginning of the last century when the normally miniscule European mole fur trade, burgeoned and nearly drove the European moles to extinction (Gorman and Stone, 1990). Fortunately, fashion has changed and the moles have once again retreated to relative obscurity.

However, despite the lack of mole breeding programs and genetic maps, it is still possible to answer some relevant questions in X-inactivation in this organism. We have already established mortal Coast mole cell lines and it should not be too difficult to create immortal cells lines through viral transformation. Once a cell lineage is established it becomes easier to conduct a whole range of experiments without the time limitations inherent in normal, mortal cells. RNA-FISH with available human and murine probes may be successful in determining whether Coast mole *Xist* coats the X chromosome, as it does in the other studied eutherians (Clemson et al., 1996, Heard et al., 1997, Duthie et al., 1999). If an association of the *Xist* RNA with the X chromosome were found, it would strengthen the results presented in this thesis, which indicate that *Xist* is probably functional in the mole. Additionally, DNA-FISH could be used to locate Coast mole *Xist* and confirm that is on the X chromosome.

It would also be possible to examine the acetylation status of histones on the Coast mole X chromosome and determine if replication-timing differences exist between the two homologues. Another avenue of interest are the LINE (L1) repeats, which are very strong candidates for the role of the booster elements that facilitate the spread of inactivation along the mouse and human X chromosome, both of which are exceptionally rich in these sequences

(Bailey et al., 1998, Lyon, 1998). Although L1 elements have been identified in a large number of eutherians and in a marsupial, the core insectivores (Euliptophyla) have not been examined for the presence of these repeats (Dorner and Paabo, 1995). One could perform FISH experiments with a conserved eutherian L1 probe to determine whether the Coast mole X chromosome is enriched for these elements as is the case in humans and mice (Lyon, 1998).

The extent and identity of genes that escape X-chromosome inactivation is another matter that could, at least theoretically, be examined in the mole. However, for this study to be feasible one must either identify X-linked Coast mole polymorphisms for each of the genes or establish a somatic cell hybrid line in which the mole X is retained in the background of another species. It may be difficult to locate polymorphisms in the mole sequences since underground mammals tend to exhibit low levels of genetic variation, presumably due to the stable and less complex underground environment (Yates and Morre, 1990). Indeed, heterozygosity values for the highly fossorial moles, such as the Coast mole, are quite low, with an average of 0.0144 (Yates and Moore, 1990). Based on a sample of 13 animals and a mean number of loci of 21.2, the heterozygosity of a *Scapanus orarius* population was estimated at 0.003 (Yates and Moore, 1990). On the other hand, the creation of somatic cell hybrids is hampered by the fact that the cells of some organisms simply fragment in a foreign background, are completely expelled or expel the chromosomes of the other species, making it impossible to maintain a stable culture.

Another very interesting question which, unfortunately may be impossible to study in the mole system, is the presence of imprinted X-inactivation. This question has only been addressed in the mouse, rat and human and results showed paternal inactivation only in the extraembryonic tissues of the two rodents (Heard et al., 1997). Since one primate and two closely related rodents hardly constitute a significant data set, it is crucial that other organisms

be examined to determine whether random inactivation in the extraembryonic tissues is a primate-specific phenomenon, possibly due to the selective advantage of psuedo-heterozygosity, or if it is the rodents that are the exception to the rule. The answer to this question would greatly illuminate the evolutionary origins of X-inactivation not only in eutherians, but also in the marsupials in which it is also the paternal chromosome that is always silenced (Cooper et al., 1993). One hypothesis for imprinted inactivation in the mouse trophoblast holds that the short gestation time and rapid development in rodents simply do not allow enough time for the erasure of epigenetic marks acquired during the transcriptional silencing of the X in spermatogenesis (Heard et al., 1997). The situation in moles is particularly interesting, not only because of their possibly basal phylogenetic position, but because, like mice, moles are small mammals with short gestation times. The average laboratory mouse lives up to three years and has a gestation period of 21 days (Green, 1966). Although moles seem to have longer life spans (one trapped European mole was thought to be seven years old!), their gestation times of approximately four weeks are comparable to those of the mouse (Gorman and Stone, 1990). If paternal inactivation were to be observed in mole extraembryonic tissues, it would support the argument that imprinted silencing is simply a consequence of rapid development. However, if this were not the case in moles, other evolutionary reasons for the situation in rodents would have to be considered. However, to determine the parental origin of the two X chromosomes it is necessary to identify X-linked polymorphisms and to genotype both parents. Unfortunately, since the animals do not breed in captivity, it is doubtful that this sort of study will be feasible in the mole in the near future. However, the question of imprinted inactivation can be addressed in other, more studied eutherian model systems such as the cow, sheep or the pig. There is considerable interest, mostly motivated by economics, in the genetics of these organisms and

genome projects are underway for the cow and pig. Furthermore, advanced cloning technology in these animals would allow the researchers to address other questions regarding X-inactivation in early mammalian development.

This study of X-chromosome inactivation in the Coast mole shows that the unique features of eutherian X-linked silencing have been conserved in the Order Insectivora. Methylation analysis of X-linked CpG islands not only shows that the Coast mole undergoes X-chromosome inactivation, but that it employs CpG methylation to achieve silencing. Furthermore, a homologue of the, at the moment, uniquely eutherian *XIST/Xist* gene has been identified in *Scapanus*. The Coast mole gene shows significant similarity to the available *Xist* sequences and is only expressed in the female animals. Despite the absence of functional studies, the combined evidence of the methylation analysis, sequence conservation and expression pattern points to a role for Coast mole *Xist* in the X-chromosome inactivation in this animal. Although the complete sequence of mole Xist will be very useful, it is unlikely that the mole system will ever be as amenable for the study of early development as other model organisms. However, while the mole does not seem suited to genetic research, the remaining Liptophylans, the hedgehogs and shrews, may be better candidates. Hedgehogs are increasingly popular as pets and can breed in captivity (Smith, 1999), while the population structure and genetic polymorphims of the shrews are well studied (Boutellier and Perrin, 2000). In addition, shrews do not seem to have problems reproducing in captivity and numerous breeding colonies have been established (Ohno et al., 1998). The hedgehog, and especially the shrew, may represent realistic candidates for the establishment of an insectivore model of X-chromosome inactivation, which I believe, would be very helpful in the elucidation of this fascinating process.

Bibliography

Adler, D.A., Rugarli, E.I., Lingenfelter, P.A., Tsuchiya, K., Poslinkski, D., Liggit, H.D., Chapman, V.M., Elliot, R.W., Ballabio, A. and Disteche, C.M. Evidence of evolutionary up-regulation of the single active X chromosome in mammals based on *ClC4* expression levels in *Mus spretus* and *Mus musculus*. *Proceedings of the National Academy of Sciences U.S.A.* 94:9244-9248, 1997.

Air, G.M, Thompson, E.O.P., Richardson, B.J. and Sharman, G.B. Amino-acid sequences of kangaroo myoglobin and haemoglobin and the date of marsupial-eutherian divergence. *Nature* 229:391-394, 1971.

Altschul, S.F., Gish, W., Miller W., Myer, E.W. and Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology*, 215:403-410, 1990.

Anderson, C.L. and Brown, C.J. Polymorphic X-chromosome inactivation of the human TIMP1 gene. *American Journal of Human Genetics* 65:699-708, 1999.

Ariel, M., Robinson, E., McCarrey, J.R. and Cedar, H. Gamete-specific methylation correlates with imprinting of the murine *Xist* gene. *Nature Genetics* 9:312-315, 1995.

Armstrong S.J., Hulten, M.A., Keohane, A.M., and Turner, B.M. Different strategies of X-inactivation in germinal and somatic cells: Histone H4 underacetylation does not mark the inactive X in the mouse male germline. *Experimental Cell Research* 230, 399-402, 1997.

Avner, P. and Heard, E. X-chromosome inactivation: Counting, choice and initiation. *Nature Reviews Genetics*, 2:59-67, 2001.

Ayoub, N., Richler, C., and Wahrman, J. *Xist* RNA is associated with the transcriptionally inactive XY body in mammalian male meiosis. *Chromosoma* 106, 1-10, 1997.

Bailey, J.A., Carrel, L., Chakravati, A. and Eichler, E. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proceedings of the National Academy of Sciences U.S.A.* 97(12):6634-6639, 2000.

Banzai M., Omoe K., Ishikawa H., Endo A. Viability, development and incidence of chromosome anomalies of preimplantation embryos from XO mice. *Cytogenetics and Cell Genetics* 70:273-277, 1995.

Baumstark, A., Akhverdyan, Schulze, A., Reisert, I., Vogel, W. and Just, W. Exclusion of *SOX9* as the testis determining factor in *Ellobius lutescens*: evidence for another testis gene besides SRY and SOX9. *Molecular Genetics and Metabolism*, 72:61-66, 2001.

Baverstock, P.R., Adams, M., Polkinhorne, R.W., and Gelder, M. A sex-linked enzyme in birds – Z chromosome conservation but no dosage compensation. *Nature* 296:763, 1982.

Beard, C., Li, E., and Jaenisch, R. Loss of methylation activates *Xist* in somatic but not in embryonic cells. *Genes and Development* 9:2325-2334, 1995.

Bouteiller, C. and Perrin, N. Individual reproductive success and effective population size in the greater white-toothed shrew *Crocidura russula*. *Proceedings of the Royal Society of London B. Biological Sciences,* 267(1444):701-705, 2000.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S. and Rastan, S. The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell,* 71:515-526, 1992.

Brown, C.J., Ballabio, A., Rupert, J.L., Lafreniere, R.G., Grompe, M., Tonlorenzi, R. and Willard, H.F. A gene from the region the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, 349:38-44, 1991.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafreniere, R.G., Xing, Y., Lawrence, J. and Willard, H.F. The Human *XIST* Gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527-542, 1992.

Brown, C.J and Willard, H.F. Molecular and genetic studies of human X chromosome inactivation. *Advances in Developmental Biology*, 2:37-72, 1993.

Brown, C.J. and Willard, H.F. The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature,* 368:154-156, 1994.

Brown, C.J., Carrel, L. and Willard, H.F. Expression of genes from the human active and inactive X chromosomes. *American Journal of Human Genetics*, 60:1333-1343, 1997.

Campbell, N.A. *Biology*. Third edition. The Benjamin/Cummings Publishing Company, 1993.

Cao, Y., Fujiwara, M., Nikaido, M., Okada, N. and Hasegawa, M. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. *Gene*, 259:149-158, 2000.

Caprara, M.G. and Nilsen, T.W. RNA: versatility in form and function. *Nature Structural Biology*, 7(10):831-833, 2000.

Carrel, L. and Willard, H.F. An assay for X inactivation based on differential methylation at the fragile X locus, *FMR1*. *American Journal of Medical Genetics*, 64:27-30, 1996a.

Carrel, L, Hunt, P.A. and Willard, H.F. Tissue and lineage-specific variation in inactive X chromosome expression of the murine *Smcx* gene. *Human Molecular Genetics* 5(9):1361-1366, 1996b.

Carrel. L., Cottle, A.A., Goglin, K.C. and Willard, H.F. A first-generation X-inactivation profile of the human X chromosome. *Proceedings of the National Academy of Sciences U.S.A.* 96(25):14440-14444, 1999.

Chen, J-L., Blasco, M. and Greider, C.W. Secondary structure of vertebrate telomerase RNA. *Cell* 100:503-514, 2000.

Chomczynski, P. and Sacchi, N. Single-step method of RNA isolation by acid guanadiniumthiocyanate-phenol-chloroform extraction. *Analytical Biochemistry*,162:156-159, 1987.

Ciccodicola, A, D'Esposito, M, Esposito, T, Gianfranco, F, Migliaccio, C, Miano, M.G., Matarazzo, M.R., Vacca, M., Franze, A., Cuccurese, M., Cocchia, M., Curci, A., Terracciano, A., Torino A., Cocchia S., Mercadente, G., Pannone, E., Archidiacono, N., Rocchi, M., Schlessinger, D. and D'Urso, M. Differentially regulated and evolved genes
in the fully sequenced Xq/Yq pseudoautosomal pseudoautosomal region. *Human Molecular Genetics* 9:395-401, 2000.

Clemson,C.M., McNeill, J.A., Willard, H.F. and Lawrence, J.B. *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear chromosome structure. *The Journal of Cell Biology*, 132:259-275, 1996.

Clemson Moulton, C., Chow, J.C., Brown, C.J. and Lawrence Bentlet, J. Stabilization and localization of *Xist* RNA are controlled by separate mechanisms and are not sufficient for X inactivation. *The Journal of Cell Biology* 142(1):13-23, 1998.

Clerc, P. and Avner, P. Role of the region 3' to *Xist* exon 6 in the counting process of X-chromosome inactivation. *Nature Genetics* 19:249-253, 1998.

Cooper, D.W, Johnston, P.G., Watson, J.M. and Graves, J.A.M. X-inactivation in marsupials and monotremes. *Seminars in Developmental Biology*, 4:117-128, 1993.

Cooper, A. and Fortey, R. Evolutionary explosions and the phylogenetic fuse. *Trends in Ecology and Evolution*, 113:151-156, 1998.

Costanzi, C. and Pehrson, J.R. Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature*, 393:599-601, 1998.

Courtier, B., Heard, E. and Avner. P. *Xce* haplotypes show modified methylation in a region of the active X chromosome lying 3' to *Xist. Proceedings of the National Academy of Sciences U.S.A.* 92:3531-3535, 1995.

Csankovszki, G., Panning, B., Bates, B., Pehrson, J.R. and Jaenisch, R. Conditional deletion of *Xist* disrupts histone macroH2A localization but not maintenance of X inactivation. *Nature Genetics*, 22:3230324, 1999.

Cutler-Allen, R., Zoghbi, H.Y., Moseley, A.B., Rosenblatt, H.M and Belmont, J.W. Methylation of *HpaII* and *HhaI* sites near the polymorphic CAG repeat in human androgen-receptor gene correlates with X chromosome inactivation. *American Journal of Human Genetics,* 51:1229-1239, 1992.

Debrand, E., Heard, E. and Avner, P. Cloning and localization of the murine *Xpct* gene: evidence for complex rearrangements during the evolution of the region around the *Xist* gene. *Genomics*, 48:296-303, 1998.

De La Fuente, R., Hahnel, A., Basrur, P.K. and King, W.A. X inactive-specific transcript (*Xist*) expression and X chromosome inactivation in the preattachment bovine embryo. *Biology of Reproduction*, 60:769-775, 1999.

Disteche, C.M. The great escape. *American Journal of Human Genetics* 60:1312-1315, 1997.

Dorner, M. and Paabo, S. Nucleotide sequence of a marsupial LINE-1 element and the evolution of placental mammals. *Molecular Biology and Evolution*, 12(5):944-948, 1995.

Duthie, S.M., Nesterova, T.B., Formstone, E.J., Keohane, A.M., Turner, B.M., Zakian, S.M. and Brockdorff, N. *Xist* RNA exhibits a banded localization on the inactive X chromosome and is excluded from autosomal material *in cis. Human Molecular Genetics*, 8:195-204, 1999.

Eggan, K., Akutsu, H., Hochedlinger, K., Rideout III, W., Tanagimachi, R. and Jaenisch, R. X-chromosome inactivation in cloned mouse embryos. *Science*, 290(5496):1578-1581, 2000.

Eichler, E.E., Kunst, C.B., Lugenbeel, K.A., Ryder, O.A., Davison, D., Warren, S.T. and Nelson, D.L. Evolution of the cryptic *FMR1* CGG repeat. *Nature Genetics*, 11:301-308, 1995.

Foster, J.W., Breman, F.E., Hampikian, G.K., Goodfellow, P.N., Sinclair, A.H., Lovell-Badges, R., Selwood, L., Renfree, M.B., Cooper, D.W. and Graves, J.A.M. Evolution of sex determination and the Y chromosome: *SRY*-related sequences in marsupials. *Nature*, 359:531-533, 1992.

Fu, Y-H, Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, Verkerk, A.J.M.H., Holde, J.J.A., Fenwick, R.G, Warren, S.T., Oostra, B.A, Nelson, D.A. and Cakey, T.C. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell*,67:1047-1058, 1991.

Fumagalli, L., Taberlet, P., Stewart, D.T., Gielly, L., Hausser, J and Vogel, P. Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, 11(2):222-235, 1999.

Gartler, S.M. and Riggs, A.D. Mammalian X-chromosome inactivation. *Annual Review of Genetics* 17, 155-190, 1983.

Gartler, S.M. and Goldman, M. Reactivation of inactive X-linked genes. *Developmental Genetics*, 15:504-514, 1994.

Gilbert, S.L and Sharp, P.A. Promoter-specific hypoacetylation of X-inactivated genes. *Proceedings of the National Academy of Sciences U.S.A.*, 96(24):13825-13830, 1999.

Gilbert, S.L, Pehrson, J.R. and Sharp, P.A. *XIST* RNA associates with specific regions of the inactive X chromatin. *The Journal of Biological Chemistry*, 275(47):36491-36494, 2000.

Gorman, M.L. and Stone, R.D. *The natural history of moles*. Comstock Publishing, Cornell University Press, New York, 1990.

Goto, T. and Monk, M. Regulation of X-chromosome inactivation in development in mice and humans. *Microbiology and Molecular Biology Reviews*, 62(2):362-378, 1998.

Goto, Y. and Takagi, N. Maternally inherited X chromosome is not inactivated in mouse blastocysts due to parental imprinting. *Chromosome Research* 7:101-109, 1999.

Graves, J.A.M., Disteche, C.M. and Toder R. Gene dosage in the evolution and function of mammalian sex chromosomes. *Cytogenetics and Cell Genetics* 80:94-103, 1998a.

Graves, J.A.M., Wakiefield, M.J. and Toder, R. The origin and evolution of the pseudoautosomal regions of human sex chromosomes. *Human Molecular Genetics*, 7(13):1191-1196, 1998b.

Green, E.L.ed. The staff of the Jackson laboratory. *Biology of the laboratory mouse.* Dover Publications, New York, 1966.

Hansen, S.R, Canfield, T.K, Stanke, A.M., Keitges, E.A. and Gartler, S.M. Reactivation of *XIST* in normal fibroblasts and a somatic cell hybrid: Abnormal localization of *XIST* RNA in hybrid cells. *Proceedings of the National Academy of Sciences U.S.A.* 95:5133-5138, 1998.

Heard, E., Clerc, P. and Avner, P. X-chromosome inactivation in mammals. *Annual Review of Genetics*, 31:571-610, 1997.

Heard, E., Mongetard, F., Arnaud, D., Chureau, C., Vourc'h, C. and Avner, P. Human *XIST* yeast artificial chromosome transgenes show partial X inactivation center function in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences U.S.A.* 96:6841-6846, 1999.

Hendrich, B.D., Brown, C.J. and Willard, H.F. Evolutionary conservation of possible functional domains of the human and murine *XIST* genes. *Human Molecular Genetics* 2(6):663-672, 1993.

Hendriksen. P.M., Hoogerbrugge, J.W., Themmen, A.P., Koken, M.H.M., Hoeijmakers, J.H.J., Oostra, B.A., Van Der Lende, T. and Grootegoed, A. Postmeiotic transcription of X and Y chromosomal genes during spermatogenesis in the mouse. *Developmental Biology* 170:730-733, 1995.

Henke, W., Herdel, K., Jung, K., Schnorr, D. and Loening, S. Betaine improves the PCR amplification of GC-rich DNA sequences, *Nucleic Acids Research* 25(19): 3957-3958, 1997.

Herzing. L.B., Romer, J.T., Horn, J.M. and Ashworth, A. *Xist* has properties of the X-chromosome inactivation centre. *Nature* 386:272-275, 1997.

Hong, Y-K., Ontiveros, S.D. and Strauss, W.M. A revision of the human *XIST* gene organization and structural comparison with mouse *Xist*. *Mammalian Genome* 11:220-224, 2000.

Hoyer-Fender, S., Costanzi, C. and Pehrson, J.R. Histone MacroH2A1.2 is concentrated in the XY-body by the early pachytene stage of spermatogenesis. *Experimental Cell Research*, 258:254-260, 2000.

Hurst, L.D. Evolutionary theories of genomic imprinting. *Genomic Imprinting*. Reik W, Surani A. (Eds.). Oxford, Oxford University Press, 1997.

Hurst, L.D. and Smith, N.G.C. Molecular evidence that *H19* mRNA is functional. *Trends in Genetics*, 4:134-135, 1999.

Jamieson, R.V., Zhou S.X., Tan S.S. X-chromosome inactivation during the development of the male urogenital ridge in the mouse. *International Journal of Developmental Biology* 41:49-55, 1997.

Jegalian, K. and Page, D.C. A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature*, 394:776-780, 1998.

Jimenez, R., Burgos, M., Sanchez, A, Sinclair, A.H., Alarcon , F.J., Marin, J.J., Ortega, E. and Diaz de la Guardia, R. Fertile females of the mole *Talpa occidentalis* are phenotypic intersexes with ovotestes. *Development*, 118:1303-1311, 1993.

Johnston, C.M., Nesterova, T.B., Formstone, E.J., Newall, A.E.T., Duthie, S.M., Sheardown, S.A. and Brockdorff, N. Develpomentally regulated *Xist* promoter switch mediates initiation of X inactivation. *Cell* 94:809-817, 1998.

Jones, P.L., Veenstra, G.J.C, Veenstra, G.J.C., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J. and Wolffe, A.P. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genetics*, 19:187191, 1998.

Juan, V, Crain, C. and Wilson, C. Evidence for evolutionarily conserved secondary structure in the *H19* tumor suppressor RNA. *Nucleic Acids Research* 28(5):1221-1227, 2000.

Just, W., Rau, W., Vogel, W., Akhverdian, M., Fredga, K., Graves, J.A.M. and Lyapunova, E. Absence of *Sry* in species of the vole *Ellobius*. *Nature Genetics*, 11:117-118, 1995.

Kaslow, D.C. and Migeon, B.R. DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: Evidence for multistep maintenance of mammalian X dosage compensation. *Proceedings of the National Academy of Sciences U.S.A.* 84:6210-6214, 1987.

Kay, G.F., Penny, G.D., Patel, D., Ashworth, A., Brockdorff, N., and Rastan, S. Expression of *Xist* during mouse development suggests a role in the initiation of X chromosome inactivation. *Cell* 72, 171-182, 1993.

Kelley, R.L. and Kuroda, M. I. The role of chromosomal RNAs in marking the X for dosage compensation. *Current Opinion in Genetics and Development*, 10:555-561, 2000a.

Kelley, R.L. and Kuroda, M. I. Noncoding RNA genes in dosage compensation and imprinting. *Cell*, 103:9-12, 2000b.

Keohane, A.M., Lavender, J.S., O'Neill, L.P. and Turner, B.M. Histone acetylation and X inactivation. *Developmental Genetics*, 22:65-73, 1998.

Kettlewell, J.R., Raymond, C.S. and Zarkower, D. Temperature-dependent expression of turtle *Dmrt1* prior to sexual differentiation. *Genesis*, 26:174-178, 2000.

Kloc, M., Spohr, G. and Etkin, L.D. Translocation of repetitive RNA sequences with the germ plasm in *Xenopus* oocytes. *Science*, 262:1712-1714, 1993.

Krettek, A., Gullberg, A. and Arnason, U. Sequence analysis of the complete mitochondrial DNA molecule of the Hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Liptophyla. *Journal of Molecular Evol*ution, 41:952-957, 1995.

Lahn, B. and Page, D.C. Functional coherence of the human Y chromosome. *Science*, 278:675-680, 1997.

Lahn, B.T. and Page, D.C. Four evolutionary strata on the human X chromosome. *Science*, 286:964-967, 1999.

Lahn, B.T., Pearson, N.M. and Jegalian, K. The human Y chromosome, in the light of evolution. *Nature Reviews Genetics*, 2:207-216, 2001.

Lee, J.T. and Jaenisch, R. Long-range *cis* effects of ectopic X-inactivation centres on a mouse autosome. *Nature* 386:275-X, 1997.

Lee, J.T. and Lu, N. Targeted mutagenesis of *Tsix* leads to nonrandom X inactivation. *Cell* 99:47-57, 1999a.

Lee, J.T., Davidow, L.S. and Warshawsky, D. *Tsix*, a gene antisense to *Xist* at the X-inactivation centre. *Nature Genetics* 21:400-404, 1999b.

Lee, J.T. Disruption of imprinted X inactivation by parent-of-origin effects at *Tsix*. *Cell* 103:17-27, 2000.

Lei, H., Oh, S.P., Okano, M., Jutterman, R., Goss, K.A., Jaenisch, R. and Li, E. *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, 122:3195-3205, 1996.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F. and Bird, A. Purification, sequence and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69:905-914, 1992.

Li, E., Bestor, T.H. and Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell,* 69:915-926, 1992.

Li, W-H., Ellsworth, D.L., Krushkal, J., Chang, B.H-J. and Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution*, 5(1):182-187, 1996.

Liu, F.R., Miyamoto, M.M., Freire, N.P., Ong, P.Q., Tennat, M.R., Young, T.S. and Gugel, K.F. Molecular and morphological supertrees for eutherian (placental) mammals. *Science,* 291:1786-1789, 2001.

Loebel, D.A.F. and Johnston, P.G. Methylation analysis of a marsupial X-linked CpG island by bisulfite genomic sequencing. *Genome Research*, 6:114-123, 1996.

Looijenga, L.H., Gillis, A.J., Verkerk, A.J., Van Putten, W.L. and Oosterhuis, J.W. Heterogeneous X-inactivation in human full-term female placentas. *American Journal of Human Genetics*, 64(5):1445-1452, 1999.

Lucchesi, J.C. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. *Current Opinion in Genetics and Development*, 8:179-184, 1998.
Luoh, S-W., Jegalian, K., Lee, A., Chen, E.Y., Ridley, A. and Page, D.C. CpG islands in human *ZFX* and *ZFY* and mouse *Zfx* genes: sequence similarities and methylation differences. *Genomics,* 29:353-363, 1995.

Lyko, F. DNA methylation learns to fly. *Trends in Genetics*, 17(4):169-172, 2001.

Lyon, M.F. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature,* 190:372-373, 1961.

Lyon, M.F. X-Chromosome inactivation: a repeat hypothesis. *Cytogenetics and Cell Genetics*, 80:133-137, 1998.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W. and Springer, M. Parallel adaptive radiations in two major clades of placental mammals. *Nature*, 409:610-614, 2001.

Marahrens, Y., Panning, B., Dausman, J., Strauss, W. and Jaenisch, R. *Xist*-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes and Development* 11:156-166, 1997.

McDonald, L.E., Paterson, C.A. and Kay, G.F. Bisulfite genomic sequencing-derived methylation profile of the *Xist* gene throughout early mouse development. *Genomics* 54:379-386, 1998.

Meller, V. Dosage compensation: making 1X equal 2X. *Trends in Cell Biology*, 10(2):54-59, 2000.

Migeon, B.R. Non-random X chromosome inactivation in mammalian cells. *Cytogenetics and Cell Genetics* 80:142-148, 1998.

Migeon, B.R., Kazi, E., Haisley-Royster, C., Hu, J., Reeves, R., Call, L., Lawler, A., Moore, C.S., Morison, H. and Jeppesen, P. Human X inactivation center induces random X chromosome inactivation in male transgenic mice. *Genomics* 59:113-121, 1999.

Migeon, B.R., Winter, H., Kazi, E., Chowdhury, A.K., Hughes, A., Haisley-Royster, C., Morison, H. and Jeppesen, P. Low-copy-number human transgene is recognized as an X inactivation center in mouse ES cells, but fails to induce *cis*-inactivation in chimeric mice. *Genomics* 71:156-162, 2001.

Mlynarczyk and Panning, B. X Inactivation: *Tsix* and *Xist* as yin and yang. *Current Biology*, 10:R899-R903, 2000.

Mouchaty, S.K., Gullberg, A., Janke, A. and Aranson, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Molecular Biology and Evolution*, 17(1):60-67, 2000.

Murphy, W.J., Eizirik, E., Johnson, w.E., Zhang, J.P., Ryder, O. and O'Brien, S.J. Molecular phylogenetics and the origin of placental mammals. *Nature*, 409:614-618, 2001.

Nagorsen, D. *Opossums, shrews and moles of British Columbia*. Vancouver, UBC Press, 1996.

Nanda, I., Zend-Ajusch, E., Shan, Z., Grutzner, F., Schartl, M., Burt, D.W., Koehler, M., Fowler, V.M., Goodwin, G., Schneider, W.J., Mizuno, S., Dechant, G., Haaf, T. and Schmid, M. Conserved synteny between the chicken Z sex chromosome and human chromosome 9 includes the male regulatory gene *DMRT1*: a comparative (re)view on avian sex determination. *Cytogenetics and Cell Genetics*, 89:67-78, 2000.

Nesterova, T.B., Duthie, S.M., Mazoruk, N.A., Isaenko, A., Rubtsova, N., Zakian, S.M. and Brockdorff, N. Comparative mapping of X chromosomes in vole species of the genus *Microtus*. *Chromosome Research*, 6:41-48, 1998.

Nesterova, T.B., Slobodyanyuk, S.Y., Elisaphenko, E.A., Shevchenko, A. I., Johnston, C., Pavlova, M.E., Rogozin, I.B., Kolesnikov, N.N., Brockdorff, N. and Zakian, S.M. Characterization of the genomic *Xist* locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Research*, 2001, in press.

Ng, H.H. and Bird, A. DNA methylation and chromatin modification. *Current Opinion in Genetics and Develpoment* 9:158-163, 1999.

Norris, D.P., Brockdorff, N. and Rastan. S. Methylation status of CpG-rich islands on active and inactive X-chromosomes. *Mammalian Genome*, 1(2):78-83, 1991.

Novacek, M.J. Mammalian phylogeny: shaking the tree. *Nature*, 356:121-125, 1992.

Ohlsson, R., Paldi, A. and Graves, J.A.M. Did genomic imprinting X chromosome inactivation arise from stochastic expression? *Trends in Genetics*, 17(3):136-141, 2001.

Ohno, S. *Sex Chromosome and Sex-linked Genes.* Monographs on endocrinology. Berlin, Heidelberg, New York. Springer-Verlag, 1967

Ohno, T., Yoshida F., Ichikawa Y., Matsuo S., Hotta N., Terada M., Tanaka S., Yamashita K, Namikawa T. and Kitoh, J. A new spontaneous animal model of NIDDM without obesity in the musk shrew. *Life Sciences*, 62(11):995-1006, 1998.

Okano, M., Bell, D.W., Daniel, A.H. and Li, E. DNA methylatransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*, 99:247-257, 1999.

O'Neill, L.P., Keohane, A.M, Lavender, J.S., McCabe, V., Heard, E., Avner, P., Brockdorff, N. and Turner, B.M. A developmental switch in H4 acetylation upstream of *Xist* plays a role in X chromosome inactivation. *The EMBO Journal*, 18(10):2897-2907, 1999.

Panning, B. and Jaenisch, R. DNA hypomethylation can activate *Xist* expression and silence X-linked genes. *Genes and Development* 10:1991-2002, 1996.

Panning, B., Dausman, J. and Jaenisch, R. X chromosome inactivation is mediated by *Xist* RNA stabilization. *Cell* 90:907-916, 1997.

Panning, B. and Jaenisch, R. RNA and the epigenetic regulation of X chromosome inactivation. *Cell*, 93:305-308, 1998.

Pask, A. and Graves, J.M. Sex chromosomes and sex-determining genes: insights from marsupials and monotremes. *Cellular and Molecular Life Sciences* 55:864-875, 1999.

Pask, A., Renfree, M.B. and Marshall Graves, J.A. The human sex-reversing *ATRX* gene has a homologue on the marsupial Y chromosome, *ATRY*: implications for the evolution of mammalian sex determination. *Proceedings of the National Academy of Sciences U.S.A.* 97(24): 13198-13202, 2000.

Pehrson, J.R. and Fuji, R.N. Evolutionary conservation of histone macroH2A subtypes and domains. *Nucleic Acids Research*, 26(12);2837-2842, 1998.

Penny, G.D., Kay, G.F, Sheardown, S.A., Rastan, S. and Brockdorff, N. Requirement for *Xist* in X chromosome inactivation. *Nature* 379:131-137, 1996.

Prissette, M., El-Maari, O., Arnaud, D., Walter, J. and Avner, P. Methylation profile of *DXPAS34* during the onset of X-inactivation. *Human Molecular Genetics*, 10(1):31-38, 2001.

Robertson, K.D. and Jones. P.A. DNA methylation: past, present and future directions. *Carcinogenesis,* 21(3):461-4667, 2000.

Robertson, K.D. and Wolffe, A.P. DNA methylation in health and disease. *Nature Reviews Genetics*, 1:11-19, 2000.

Sado, T., Fenner, M.H., Tan, S-S., Tam, P., Shioda, T. and Li, E. X Inactivation in the mouse embryo deficient for *Dnmt1*: distinct effect of hypomethylation on imprinted and random X inactivation. *Developmental Biology*, 225:294-303, 2000.

Sanchez, A., Bullejos, M., Burgos, M., Hera, C., Stamatopoulos, C., Diaz de la Guardia, R., and Jimenez, R. Females of four mole species of genus *Talpa* (Insectivora, Mammalia) are true hermaphrodites with ovotestes. *Molecular Reproduction and Development,* 44:289-294, 1996a.

Sanchez, A., Bullejos, M., Burgos, M., Hera, C., Jimenez, R. and Diaz de la Guardia, R. High sequence identity between the SRY HMG box from humans and insectivores. *Mammalian Genome,* 7:536-538, 1996b.

Sheardown, S., Norris, D., Fisher, A. and Brockdorff, N. The mouse *Smcx* gene exhibits developmental and tissue specific variation in degree of escape from X inactivation. *Human Molecular Genetics* 5(9):1355-1360, 1996.

Sinclair, A.H, Berta, P., Palmer, M.S., Hawkins, J.R., Grffiths, B.L., Smith, M.J., Foster, M.J., Frischauf, A.M., Lovell-Badge, R. And Goodfellow, P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240-244, 1990.

Smith, A. J. Husbandry and nutrition of hedgehogs. *Veterinary Clinics of North America. Exotic Animal Practice*, 1:127-141, 1999.

Solari, A. *Sex Chromosomes and Sex Determination in Vertebrates*. Boca Raton, Ann Arbor, London, Tokyo. CRC Press, 1993.

Solter, D. and Wei, G. Ends *Xist*, but where are the beginings? *Genes and Development* 11:153-155, 1997.

Springer, M.S., De Bry, R.W., Douady, C., Amrine, H.M., Madsen, O., de Jong, W. and Stanhope, M.J. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. *Molecular Biology and Evolution*, 18(2):132-143, 2001.

Stanhope, M.J., Waddel, V.G., Madsen, O., De Jong, W., Hedges, S.B., Cleven, G.C., Kao, D. and Springer, M.S. Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proceedings of the National Academy of Sciences U.S.A.*, 95:9967-9972, 1998.

Strachan, T. and Read, A.P. *Human Molecular Genetics 2*. BIOS Scientific Publishers, 1999.

Takagi, N. And Sasaki, M. Preferential inactivation of the paternally derived X-chromosome in the extraembryonic membranes of the mouse. *Nature* 256:640-642, 1975.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22(22):4673-4680, 1994.

Tinker, A.V. and Brown, C.J. Induction of *XIST* expression from the human active X chromosome in mouse/human somatic cell hybrids by DNA demethylation. *Nucleic Acids Research* 26(12):2935-2940, 1998.

Uberbacher, E.C. and Mural, R.J. Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proceedings of the National Academy of Sciences U.S.A.* 88:11261-11265, 1991

Vogel, W., Jainta, S., Rau, W., Geerkens, C., Baumstark, A., Correa-Cerro, L.-S., Ebenhoch, C. and Just, W. Sex determination in *Ellobius lutescens*: the story of an enigma. *Cytogenetics and Cell Genetics* 80:214-221, 1998.

Wachtel, S.S. X-linked sex-reversing genes. *Cytogenetics and Cell Genetics*, 80:222-225, 1998.

Wake, N., Tagaki, N. and Sasaki, M. Non-random inactivation of X chromosome in the rat yolk sac. *Nature*, 262:580-581, 1976.

Wakefield, M.J., Keohane, A.M., Turner, B.M. and Graves, J.A.M. Histone underacetylation is an ancient component of mammalian X chromosome inactivation. *Proceedings of the National Academy of Sciences U.S.A.*, 94:9665-9668, 1997.

Warshawsky, D., Stavropoulos, N. and Lee, J.T. Further examination of the *Xist* promoter-switch hypothesis in X inactivation: evidence against the existence and function of a Po promoter. *Proceedings of the National Academy of Sciences U.S.A.* 96(25):14424-14429, 1999.

Watson, J.M., Spencer, J.A., Riggs, A.D. and Marshall Graves, J.A. The X chromosome of monotremes shares a highly conserved region with the eutherian and marsupial X chromosomes despite the absence of X chromosome inactivation. *Proceedings of the National Academy of Sciences U.S.A.* 87:7125-7129, 1990.

Watson, J.M., Spencer, J.A., Riggs, A.D. and Graves, J.A.M. Sex chromosome evolution: Platypus gene mapping suggests that part of the human X chromosome was originally autosomal. *Proceedings of the National Academy of Sciences U.S.A.*, 88:11256-11260, 1991.

Waugh O'Neill, R.J., O'Neill, M. and Graves, J.A.M. Undermethylation associated with retroelement activation and chromosome remodelling in an interspecific mammalian hybrid. *Nature*, 393:68-72, 1998.

Wen-Hsiung. L., Ellsworth, D.L., Krushkal, J., Chang, B. H-J. and Hewett-Emmett, D. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Molecular Phylogenetics and Evolution* 5(1):182-187, 1996.

Whitworth, D.J., Licht, P, Racey, P.A. and Glickman, S.E. Testis-like steroidogenesis in the ovotestis of the European mole, *Talpa europea. Biology of Reproduction*, 60:413-418, 1999.

Wutz, A. and Jaenisch, R. A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Molecular Cell* 5:695-705, 2000.

Wrigley, J.M. and Graves, J.A.M. Sex chromosome homology and incomplete, tissue-specific X-inactivation suggest that monotremes represent an intermediate stage of mammalian sex chromosome evolution. *Journal of Heredity* 79:115-118, 1988.

Yates, T.L. and Moore, D.W. Speciation and evolution in the family Talpidae (Mammalia: Insectivora), *Progress in Clinical and Biological Research*, 335:1-22, 1990.

Zarkower, D. Establishing sexual dimorphism: conservation amidst diversity? *Nature Reviews Genetics*, 2:175-185, 2001.

Zuccotti, M. and Monk, M. Methylation of the mouse *Xist* gene in sperm and eggs correlates with imprinted *Xist* expression and paternal X-inactivation. *Nature Genetics* 9:316-320, 1995.

Appendix 1: *XIST/Xist* sequences used in the creation of Figures 8 and 15

Primate

Human	M97168
Rhesus monkey	L10733, AF140294 (EST), BE483406 (EST)
Chimpanzee	L10729
Howler monkey	L10731
African green monkey	L10723

Rodent

Mouse	L04961
Vole (*M. rossiameridionalis*)	AJ310130
Vole (*M. arvalis*)	AJ310129
Vole (*M. transcaspicus*)	AJ310127
Rat (ESTs)	BF419051, AI145414.1, BF415216

Others

Horse	U50911
Rabbit	U50910
Cat	AF197966
Cow	AF104906, AV617038 (EST)
Pig	BF702572