

Research

Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate *XIST* locus

Julie E. Horvath,^{1,2,6} Christina B. Sheedy,¹ Stephanie L. Merrett,¹ Abdoulaye Banire Diallo,³ David L. Swofford,^{1,4} NISC Comparative Sequencing Program,⁵ Eric D. Green,⁵ and Huntington F. Willard^{1,4}

¹Duke Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA; ²Department of Evolutionary Anthropology, Duke University, Durham, North Carolina 27708, USA; ³Département d'informatique, Université du Québec À Montréal, Montréal H3C 3P8, Canada; ⁴Department of Biology, Duke University, Durham, North Carolina 27708, USA; ⁵Genome Technology Branch and NIH Intramural Sequencing Center (NISC), National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

Here we provide a detailed comparative analysis across the candidate X-inactivation Center (XIC) region and the *XIST* locus in the genomes of six primates and three mammalian outgroup species. Since lemurs and other strepsirrhine primates represent the sister lineage to all other primates, this analysis focuses on lemurs to reconstruct the ancestral primate sequences and to gain insight into the evolution of this region and the genes within it. This comparative evolutionary genomics approach reveals significant expansion in genomic size across the XIC region in higher primates, with minimal size alterations across the *XIST* locus itself. Reconstructed primate ancestral XIC sequences show that the most dramatic changes during the past 80 million years occurred between the ancestral primate and the lineage leading to Old World monkeys. In contrast, the *XIST* locus compared between human and the primate ancestor does not indicate any dramatic changes to exons or *XIST*-specific repeats; rather, evolution of this locus reflects small incremental changes in overall sequence identity and short repeat insertions. While this comparative analysis reinforces that the region around *XIST* has been subject to significant genomic change, even among primates, our data suggest that evolution of the *XIST* sequences themselves represents only small lineage-specific changes across the past 80 million years.

[Supplemental material is available for this article. The sequence data from this study have been submitted to GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) under accession nos. AC204188, AC203493, AC204810, AC203729, and FJ156094-96.]

Coupled with new sequencing technologies that allow broader sampling from the evolutionary tree, comparative genomics is a powerful approach for understanding evolutionary changes in genome architecture and their potential implications for genome function. Multispecies sequence comparisons among placental mammals have allowed identification of lineage-specific elements (Boffelli et al. 2003) and rapidly evolving gene families (Cheng et al. 2005). Chromosome-level comparative studies in mammalian genomes have allowed reconstruction of ancestral mammalian karyotypes (Murphy et al. 2005; for review, see Ferguson-Smith and Trifonov 2007) and have revealed a much more recent origin of the sex chromosomes than previously thought (Veyrunes et al. 2008). This latter finding is of particular significance since the process of dosage compensation (equalizing gene expression on male and female sex chromosomes) is considered to be conserved among mammals, albeit with evident genomic and mechanistic differences (for review, see Okamoto and Heard 2009).

The dosage compensation mechanism that evolved in eutherian mammals is termed “X chromosome inactivation” and is achieved by randomly choosing to transcriptionally silence one of

the two X chromosomes in each female cell during early development (Lyon 1961; Erwin and Lee 2008). Comparative analysis of mammalian X inactivation will offer clues into the evolution of dosage compensation and epigenetic silencing and provide potential insight into the genomic basis for differences in various mechanisms of X inactivation, both among placental mammals and between placental and nonplacental mammals.

Studies in both humans and mice have indicated a region called the “X-Inactivation Center” (XIC/Xic) that is crucial for X inactivation (Fig. 1; Therman et al. 1974; Rastan 1983; Brown et al. 1991b). The candidate XIC/Xic is involved in the initiation and propagation of X inactivation and has therefore been the focus of many comparative studies. In humans, the XIC region was originally localized by analysis of cell lines from patients with X-chromosome abnormalities (Brown et al. 1991b; Lafreniere et al. 1993; Leppig et al. 1993). Efforts to refine the mouse Xic region have relied on analysis of both naturally occurring and engineered variants, although no definitive region has been clearly and unequivocally agreed on (Rastan 1983; Heard et al. 1996, 1999; Lee et al. 1996, 1999b; Matsuura et al. 1996; Herzog et al. 1997). This candidate XIC/Xic region contains several protein-coding and noncoding RNAs, although most do not appear to play a role in X inactivation. As initially described in the human XIC, the critical effector gene for X inactivation is the X-inactivation-specific transcript (*XIST*) gene, the product of which is a long noncoding

⁶Corresponding author.

E-mail juliann.horvath@duke.edu.

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.111849.110>.

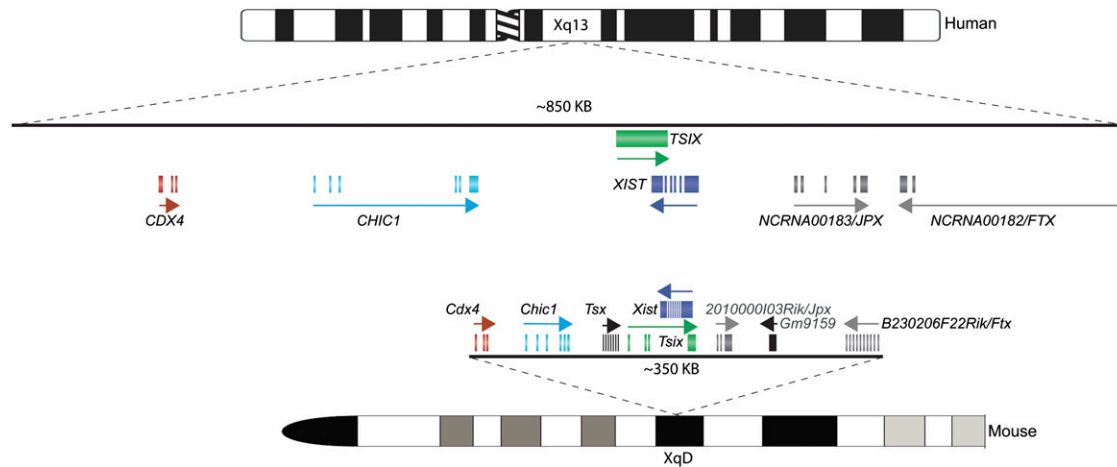


Figure 1. Candidate XIC region in human and mouse. A schematic of the gene content and organization across the candidate X-Inactivation Center (XIC) of human (72.5–73.4 Mb in hg19) and the orthologous region of mouse is shown. RefSeq genes orthologous between human and mouse (gray and colored boxes). (Black) Genes that are not orthologous. (Arrows) Indicate the direction of transcription of each gene. The human *NCRNA00183* is also referred to as *JPX* or *ENOX*, while mouse RefSeq *2010000103Rik* corresponds to *Jpx*. Mouse RefSeq *2010000103Rik* is now referred to as *Enox*. The human *NCRNA00182* is also referred to as *FTX*, while in mouse RefSeq *B230206F22Rik* corresponds to *Ftx*. Human *NCRNA00182* is now referred to as *MIR374AHG*.

RNA that is expressed exclusively from the inactive X chromosome (Brown et al. 1991a). Subsequent studies of the orthologous gene in mouse (Borsani et al. 1991; Brockdorff et al. 1991) have shown that *Xist* is required for X inactivation to occur (Penny et al. 1996).

Comparative studies among human, mouse, cow, and vole (Hendrich et al. 1997; Nesterova and Slobodyanyuk 2001; Chureau et al. 2002; Yen et al. 2007) have suggested that, although the general underlying mechanism of X inactivation in these species appears to be maintained, their different evolutionary paths have allowed for lineage-specific changes that can help elucidate sequence features that are critical for X inactivation. These comparative studies highlight a different *XIST/Xist* gene structure, different frequencies of interspersed repeat elements, and differential inactive X (Xi) chromosome chromatin that is formed via histone variants and histone modifications (for review, see Chadwick and Willard 2003). Comparative studies in marsupials, monotremes, and chicken failed to identify an *XIST* ortholog but instead identified a protein-coding gene, *LnX3*, which has several exons with identifiable homology with *XIST* (Duret et al. 2006; Hore et al. 2007a). This suggests that *XIST/Xist* evolved as a key player in placental mammalian X inactivation only in the last 175 million years since the divergence of Metatheria and Eutheria (Woodburne et al. 2003).

Since the XIC region has been disrupted in marsupial and monotreme genomes (Hore et al. 2007b) and numerous differences have been identified between human and mouse X inactivation, we have used primate comparative genomics to get a better understanding of the candidate XIC in multiple primate lineages. In the primate evolutionary tree, lemurs lie at a key position for addressing aspects of ancestral primate X-chromosome organization and may shed light on aspects of X inactivation, offering hypotheses about functionally relevant regions that are independently maintained on diverse primate lineages. The ancestral lineages leading to humans and lemurs diverged more than 80 million years ago (Mya) (Murphy et al. 2007; Horvath et al. 2008; Perelman et al. 2011), allowing for enough sequence divergence to distinguish functionally conserved regions from regions conserved due to short divergence time. Here, we have focused on genomic analyses of the candidate XIC in two lemur species to elucidate

both gene content and order within the presumptive lemur XIC and to determine the structure of the lemur *XIST* genes. Finally, we have used these comparative sequences to reconstruct the ancestral primate XIC region and *XIST* gene, representing a model for the then newly evolved *XIST* gene in genomes of placental mammals some 175 Mya.

Results

Sequence changes across the candidate primate XIC region

We focused our studies on two lemur species with the available bacterial artificial chromosome (BAC) library and cell line resources (Horvath and Willard 2007), the black lemur (*Eulemur macaco macaco*) and the ring-tailed lemur (*Lemur catta*). BAC libraries from both the black lemur and the ring-tailed lemur were screened using probes from the candidate XIC region (Supplemental Fig. 1). All positive BACs were characterized and aligned by STS content mapping and BAC end sequencing. Two overlapping BACs were selected from each species for sequence analysis at the NIH Intramural Sequencing Center (NISC). Fluorescence in situ hybridization was used to verify that these BACs mapped specifically to the X chromosome in both lemur species (data not shown). The black lemur sequence from the overlapping BACs spans 335 kb, while the orthologous ring-tailed lemur sequence spans 288 kb. Both lemur XIC sequences are collinear to the human XIC (Supplemental Fig. 2), although the lemur BAC sequences do not span the entire region homologous to the human XIC. The black lemur BAC sequences span from *CDX4* exon 2 through the first two exons of the *NCRNA00183* gene (also known as *JPX/ENOX*). The ring-tailed lemur sequences encompass all exons of *CDX4* and continue through the first two exons of the *NCRNA00183* gene. All subsequent comparative analyses, therefore, include the region from *CDX4* through the first two exons of the *NCRNA00183* gene to encompass the lemur sequences.

Comparative analyses across this region indicate lineage-specific insertions and deletions (Fig. 2A; Supplemental Fig. 3), as evidenced by a larger size in the human, chimpanzee, orangutan, and macaque genomes (Fig. 3A). Sequences corresponding to

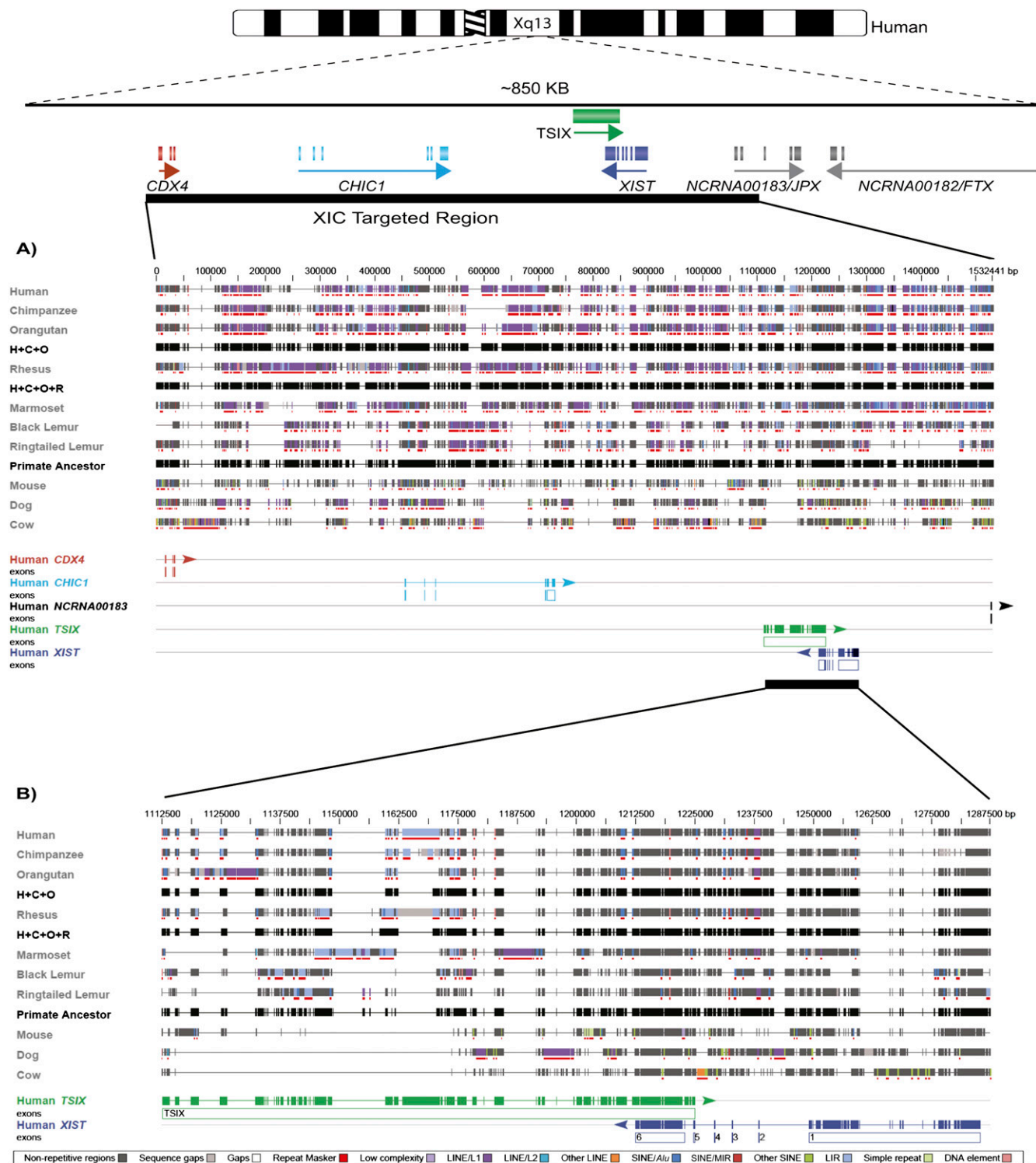


Figure 2. Repeat content and alignment across the XIC. (A) The candidate XIC region is shown along the *top* with a horizontal black bar indicating the smaller XIC region targeted in this study. Aligned sequence for each species is color-coded based on DNA sequence content, with nonrepetitive sequence indicated by black shading and different repeats color-coded according to the legend along the *bottom*. (Red bars *below* the sequence) Repetitive regions identified by RepeatMasker. Three ancestrally reconstructed sequences (H+C+O, H+C+O+R, and Primate Ancestor) are indicated for comparison of which regions have been gained or lost throughout primate evolution. *Below* the aligned sequences are the exonic regions annotated based on the human gene structures. The two exons of *JPX* occupy such a small space that they do not resolve as separate entities in this overview. (B) The aligned region expanded for *TSIX* and *XIST* is shown.

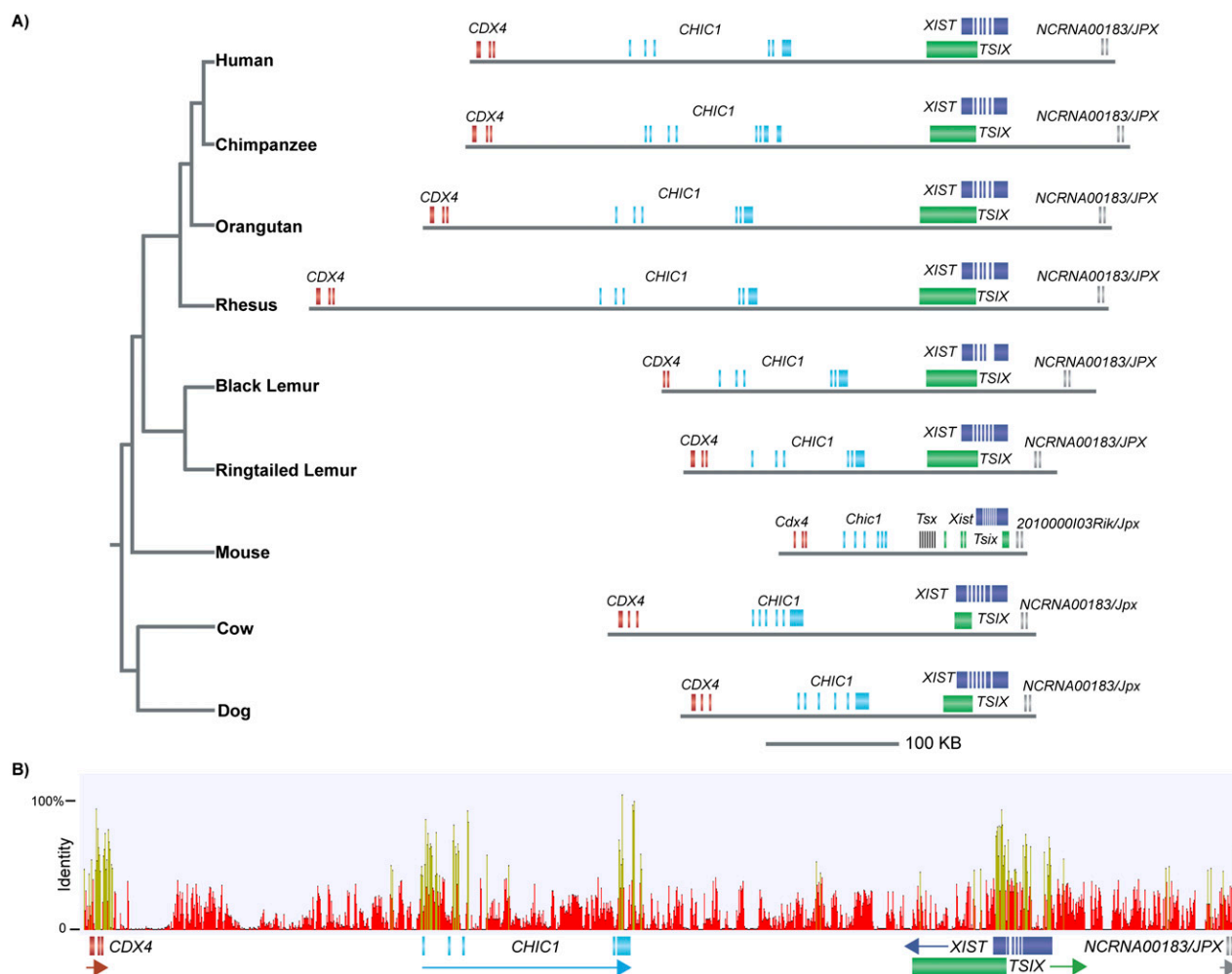


Figure 3. Candidate XIC genomic content. (A, left) The general phylogeny of the mammalian species included in this study (Murphy et al. 2007; Horvath et al. 2008; Perelman et al. 2011). Candidate XIC sequences were downloaded from UCSC for all species (except the lemurs). The lemur sequences were obtained from overlapping BACs and do not span the entire human candidate XIC region. Note that the ring-tailed lemur sequence begins at exon 1 of *CDX4*, while the black lemur sequence does not begin until exon 2 of the *CDX4* gene. Both lemur sequences extend through the first two exons of *NCRNA00183 (JPX/ENOX)*. All sequences were aligned using MLAGAN (see Methods). (Horizontal gray lines) Relative sizes of the candidate XIC in all species. (Colored boxes) DNA sequences orthologous to the annotated human gene structure. In dog and cow, the human *TSIX* sequence was used for annotation since there is no *TSIX* annotation in these species. The marmoset is not shown in this figure because the genome assembly is not complete across this region. Mouse *2010000103Rik* is now referred to as *Enox*. (B) Sequence identity plot from zero to 100% identity across the candidate XIC region for the multispecies alignment in A. (Yellow peaks) Regions of higher identity relative to red peaks or regions with no peaks. (Colored boxes) The relative positions of exons.

RefSeq genes *CDX4*, *CHIC1*, *TSIX*, *XIST*, and *NCRNA00183* are conserved in all species compared, although there are no data indicating that these are genes in the nonhuman primates, dog, or cow. Previous directed studies identified conservation of *Tsx* (Chureau et al. 2002) sequences between human and mouse. Our global alignment across the candidate XIC shows sequence alignment across some *Tsx* exons in humans and several primates, but only minimal alignment across exon 1 in the lemurs (Supplemental Fig. 3). There is no evidence to suggest that *Tsx* is a functional region in humans or any of the other primates.

Although the overall span of the XIC region is different between species (Fig. 3A), the expansion or contraction of the region is not simply the result of insertions or deletions of known repeated DNA families (Fig. 2A,B; Supplemental Fig. 3; Table 1). Rather, there are stretches of sequence unique to each species scattered throughout. Interestingly, in the rhesus macaque genome, >94 kb of sequence is inserted between *CDX4* and *CHIC1*

relative to the other primates. While a small amount of this sequence is unique to the rhesus genome (4%) or is composed of small gaps of N's (10%), the vast majority is composed of repetitive elements (86%) specific to the rhesus macaque. Not surprisingly, an identity plot showing regions of high conservation among all species indicates that the most conserved regions are exons of genes (Fig. 3B). There is also a high level of conservation over the exons of the *XIST* locus (Fig. 3B). There are small peaks of conservation outside of coding regions, the majority of which are unique sequences littered with a few conserved repetitive elements.

Reconstruction of the ancestral primate XIC

To gain further insight into the evolution of the XIC region, we used a maximum likelihood approach of ancestral reconstruction across the candidate region, using previously described methods (Blanchette et al. 2004, 2008; Diallo et al. 2007). The confidence

Table 1. Interspersed repeat content within candidate X-inactivation center

| | Human | Chimpanzee | Orangutan | Rhesus | Marmoset | Black lemur | Ring-tailed lemur | Ancestral primate | Mouse | Dog | Cow |
|--------------|-------|------------|-----------|--------|----------|-------------|-------------------|-------------------|-------|-------|-------|
| % <i>Alu</i> | 8.0 | 7.0 | 7.0 | 7.2 | 8.8 | 5.1 | 3.8 | 5.4 | 9.9 | 0.0 | 0.0 |
| %MIR | 1.2 | 1.2 | 1.0 | 0.9 | 0.9 | 1.3 | 2.1 | 2.2 | 0.6 | 2.2 | 1.2 |
| %Other SINE | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 8.9 | 6.6 | 13.7 |
| %LINE1 | 51.1 | 42.8 | 49.5 | 48.6 | 41.4 | 40.8 | 39.6 | 40.1 | 8.2 | 47.8 | 29.4 |
| %LINE2 | 1.1 | 1.0 | 0.9 | 0.8 | 1.0 | 1.0 | 1.3 | 1.8 | 0.3 | 1.5 | 1.1 |
| %Other LINE | 0.3 | 0.2 | 0.0 | 0.3 | 0.1 | 0.1 | 0.0 | 0.2 | 0.1 | 0.1 | 10.8 |
| %LTR | 10.8 | 9.4 | 9.9 | 12.3 | 10.2 | 9.9 | 9.8 | 6.8 | 4.5 | 2.5 | 5.3 |
| %DNA element | 2.3 | 2.3 | 2.2 | 1.7 | 1.6 | 2.8 | 2.7 | 2.7 | 0.4 | 1.1 | 0.6 |
| %Simple rpt | 0.5 | 0.5 | 0.6 | 0.5 | 0.4 | 0.4 | 0.5 | 0.4 | 3.1 | 0.9 | 0.5 |
| %Low complex | 0.1 | 0.1 | 0.1 | 0.1 | 0.2 | 0.4 | 0.2 | 0.2 | 1.0 | 0.3 | 0.4 |
| Size (kb) | 503.7 | 512.4 | 529.8 | 624.1 | 501.3 | 312.5 | 265.5 | 343.5 | 184.1 | 286.6 | 348.2 |
| %GC | 39.1 | 38.4 | 39.1 | 39.1 | 39.5 | 38.4 | 38.4 | 38.4 | 41.6 | 38.4 | 39.2 |

level for the presence or absence of bases in the reconstructed XIC ancestors is >80% for >90% of the bases (Supplemental Fig. 4; Diallo et al. 2010). Reconstructions of the putative human/chimpanzee/orangutan (H+C+O) ancestral XIC region indicate little overall change in size or gene content and arrangement between the extant primate species and their putative ancestor (Fig. 4). Reconstructions of the human/chimpanzee/orangutan/rhesus (H+C+O+R) ancestor are very similar, but, as noted above (Fig. 3A), indicate large lineage-specific events along the lineage leading to rhesus macaques (Fig. 4). Reconstruction of the ancestral primate candidate XIC region suggests a size comparable to the more compact current lemur XIC regions. Comparison of the various ancestral and extant XIC regions implies that significant expansion of the XIC region occurred along the H+C+O+R lineage at some point between 25 and 90 Mya (Goodman et al. 1998; Murphy et al. 2007; Horvath et al. 2008; Perelman et al. 2011). In contrast to regions such as the *CHIC1/Chic1* gene (Fig. 3A), the relative size of the *XIST/Xist* locus has remained stable across all species.

Comparative *XIST* structure

Reconstruction of the ancestral *XIST* locus

The ancestral primate *XIST* locus was reconstructed using comparisons of sequences from all primate species explored in this study, as well as the three outgroup species (mouse, dog, and cow). The confidence of the ancestral *XIST* reconstructions is quite high (>98% confidence for 95% of the positions in terms of structural content, and >80% confidence for >70% of the positions in terms of nucleotide composition) (Supplemental Fig. 5). The overall size of the inferred ancestral primate *XIST* locus (31,689 bp) is very similar to the size of the current human gene (32,094 bp). The total percentage of interspersed repeats, however, is lower in the ancestor, with the largest difference accounted for by the number of *Alu* elements (Table 2). *XIST*-specific repeats (Brockdorff et al. 1992; Brown et al. 1992; Nesterova and Slobodyanyuk 2001; Wutz et al. 2002; Hore et al. 2007b; Yen et al. 2007; Elisaphenko et al. 2008), denoted by a letter and corresponding color in Figure 5,

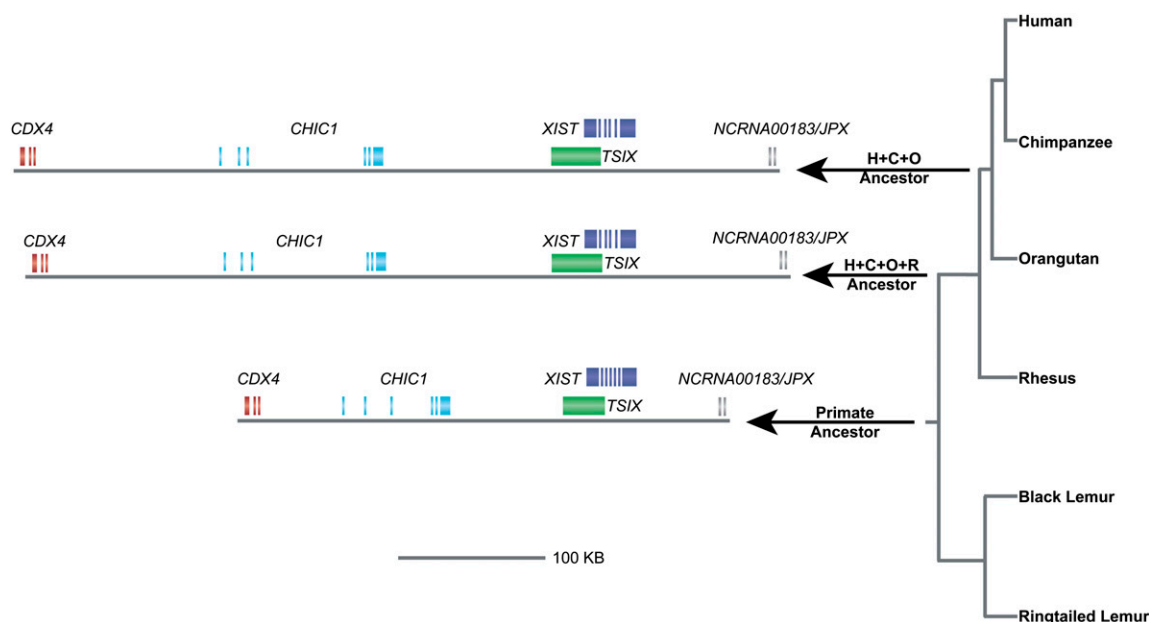


Figure 4. Ancestral reconstruction of candidate XIC region. The general phylogeny of the primates is shown along the *right* side. The approximate size and gene content of the human/chimpanzee/orangutan (H+C+O) ancestor, the human/chimpanzee/orangutan/rhesus (H+C+O+R) ancestor, and the primate ancestor, as determined through multispecies alignment reconstructions, is shown. The size of the primate ancestral XIC region is much smaller than the other ancestral sequences. This is due to repetitive and unique sequence insertions on the haplorrhine ancestral lineage after the divergence of the strepsirrhine primates.

Table 2. Interspersed repeat content across *XIST/Xist* locus

| | Human | Chimpanzee | Orangutan | Rhesus | Marmoset | Black lemur | Ring-tailed lemur | Ancestral primate | Mouse | Dog | Cow |
|----------------|--------|------------|-----------|--------|----------|-------------|-------------------|-------------------|--------|--------|--------|
| % <i>Alu</i> | 4.0 | 4.1 | 3.1 | 3.8 | 3.6 | 8.5 | 5.2 | 1.1 | 2.3 | 0.0 | 0.0 |
| %MIR | 1.4 | 1.4 | 0.5 | 1.6 | 0.6 | 1.8 | 1.8 | 3.0 | 0.3 | 0.8 | 0.9 |
| %Other SINE | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.3 | 4.5 | 6.3 |
| %LINE1 | 4.3 | 4.4 | 4.1 | 4.0 | 3.9 | 0.2 | 3.4 | 4.2 | 0.0 | 8.8 | 0.2 |
| %LINE2 | 0.9 | 0.7 | 0.8 | 1.0 | 0.3 | 0.4 | 0.4 | 1.0 | 0.0 | 0.2 | 0.3 |
| %Other LINE | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 7.4 |
| %LTR | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 |
| %DNA element | 0.5 | 0.6 | 0.5 | 0.5 | 0.4 | 0.6 | 0.6 | 0.6 | 0.0 | 0.4 | 0.0 |
| %Simple rpt | 0.7 | 0.4 | 0.8 | 0.7 | 0.3 | 1.0 | 1.2 | 1.7 | 1.6 | 1.4 | 0.8 |
| %Low complex | 0.5 | 0.3 | 0.3 | 0.3 | 0.2 | 0.1 | 0.1 | 0.2 | 2.1 | 0.2 | 0.4 |
| Total % masked | 13.8 | 13.2 | 11.6 | 13.2 | 12.0 | 13.8 | 13.7 | 11.8 | 9.6 | 17.7 | 16.5 |
| Size (bp) | 32,094 | 31,155 | 31,255 | 32,487 | 32,872 | 27,668 | 30,145 | 31,689 | 22,884 | 37,018 | 36,752 |
| %GC | 39.8 | 39.4 | 39.7 | 40.0 | 39.9 | 41.0 | 41.5 | 40.0 | 41.3 | 39.0 | 40.6 |

are similar between the ancestral primate and human forms (Table 3). This is in sharp contrast to comparisons of the human and mouse *XIST/Xist* genes, where the size is significantly different (32 kb in human vs. 22 kb in mouse), *XIST/Xist* repeat and interspersed repeat content are variable, and exon/intron structure is different (Chureau et al. 2002).

XIST-specific repeats

Given that the ancestral primate sequence did not indicate dramatic changes across the *XIST* locus, we sought to understand how different regions of *XIST* have evolved across primate genomes. We focused our attention on two *XIST*-specific repeats, the A repeat (Hendrich et al. 1997), which is highly conserved (Yen et al. 2007); and the D repeat (Yen et al. 2007), which varies between species.

The *XIST* A repeat is located at the 5' end of *XIST* exon 1 (see Fig. 5) and has been identified in almost all species studied (Brown et al. 1992; Brown and Baldry 1996; Hendrich et al. 1997; Nesterova and Slobodyanyuk 2001; Brockdorff 2002; Wutz et al. 2002; Hore et al. 2007b; Yen et al. 2007; Elisaphenko et al. 2008; Maenner et al. 2010). The A repeat, which may have been derived from an endogenous retrovirus (Elisaphenko et al. 2008), is critical for gene silencing and is essential for X inactivation (Zhao et al. 2008), but not essential for recruitment of epigenetic marks associated with X inactivation (Wutz et al. 2002). The general structure of the A repeat is a series of 42–50-bp monomer repeats separated by a spacer, which is followed by another series of repeated monomers. This general structure is observed in all primate genomes examined in this study, as well as the nonprimate outgroup species (Supplemental Fig. 6). Alignment of the A-repeat region in all species indicates significant sequence changes in mouse, dog, and cow when compared to human (Supplemental Table 1).

The D repeat is also found within *XIST* exon 1 but is not present in all species (Supplemental Figs. 7, 8). The D-repeat region is composed of a longer 290-bp monomer and is variable in size among primates and the ancestor (Supplemental Table 2). Alignment of the consensus D-repeat monomer sequences (generated via Tandem Repeats Finder) in primates revealed that the percent identity between human and other primate sequences did not always correspond to evolutionary distance from humans, perhaps indicating concerted evolution of the repeats in some or all lineages (Supplemental Table 2). Consistent with this, the D-repeat region is much larger in the cow and dog than human (Yen et al. 2007), and sequence alignments show little conservation.

Global sequence alignments using the program VISTA across all species in this study (seven primates and mouse, dog, and cow)

identified three regions that were highly conserved by RankVISTA (significance P -values $< 1 \times 10^6$) (Mayor et al. 2000; Frazer et al. 2004) in all species. One region identified (CNS1) corresponds to *XIST/Xist* exon 4, which is one of the exons believed to be derived from the chicken *Lmx3* gene (Duret et al. 2006; Elisaphenko et al. 2008) and is conserved in all species analyzed here (Fig. 6A). A second highly conserved region (CNS2) spans 240 bp across the end of exon 1, including a portion of the first intron (Fig. 6B). This region has not been shown to be functional or even highly conserved in past analyses and warrants further study. The third region (CNS3) (Fig. 6C) overlaps the A-repeat region, which is not surprising given its requirement for *XIST/Xist* function (Zhao et al. 2008).

Expression of *XIST* RNA

While the current availability of genomic resources restricted full analysis to the black and ring-tailed lemur genomes (Horvath and Willard 2007), we conducted limited analysis of other lemur species. There are five lemur taxonomic families (Daubentoniidae, Lepilemuridae, Cheirogaleidae, Lemuridae, and Indriidae), and cell lines were available for species from four of the five. To verify *XIST* expression in lemurs, male and female fibroblast cells were harvested, and RNA was isolated for cDNA characterization. Conserved primers specific to three regions across *XIST* (portions of exon 1 and exon 6,

Table 3. *XIST* regions compared between human and putative ancestral primates

| Region | Human size (bp) | Ancestral primate size (bp) | Percent identity |
|-----------|-----------------|-----------------------------|------------------|
| A repeat | 389 | 409 | 90.2 |
| B repeat | 117 | 133 | 80.5 |
| Bh repeat | 93 | 46 | 36.6 |
| C repeat | 46 | 46 | 89.1 |
| D repeat | 3776 | 2732 | 64.2 |
| E repeat | 1466 | 1556 | 88.0 |
| F repeat | 100 | 100 | 98.0 |
| CNS1 | 194 | 199 | 96.0 |
| CNS2 | 234 | 233 | 94.4 |
| CNS3 | 535 | 556 | 91.2 |
| Exon 1 | 11,372 | 10,833 | 78.6 |
| Exon 2 | 64 | 64 | 95.3 |
| Exon 3 | 137 | 138 | 89.1 |
| Exon 4 | 209 | 209 | 96.7 |
| Exon 5 | 164 | 169 | 89.9 |
| Exon 6 | 7325 | 7693 | 88.4 |

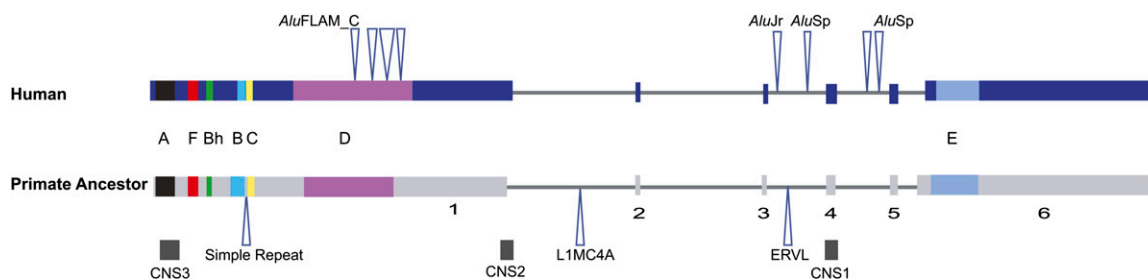


Figure 5. Reconstruction of the ancestral primate *XIST* locus and comparison of multispecies conserved regions. Horizontal lines represent the *XIST* locus in human and the reconstructed primate ancestor. Human exons (dark blue); ancestral primate DNA corresponding to those exons (gray). Each *XIST*-specific repeat (A, B, Bh, C, D, E, F) is color-coded and shown with approximate size and location. (Dark gray boxes) Approximate locations of conserved noncoding sequences (CNS). See Figure 6 for more details about each CNS. (Open triangles) Human and ancestrally reconstructed primate-specific insertions. (Triangles without a label) Insertion of nonrepetitive sequence.

and the region spanning from exons 1 to 6) verified expression solely from female cell lines from black lemur (*Lemuridae*), ring-tailed lemur (*Lemuridae*), Coquerel's sifaka (*Indriidae*), and aye-aye

(*Daubentoniidae*). Although no male mouse lemur cell lines were available for study, female mouse lemur (*Cheirogaleidae*) cDNA verified expression using the three primer sets (data not shown).

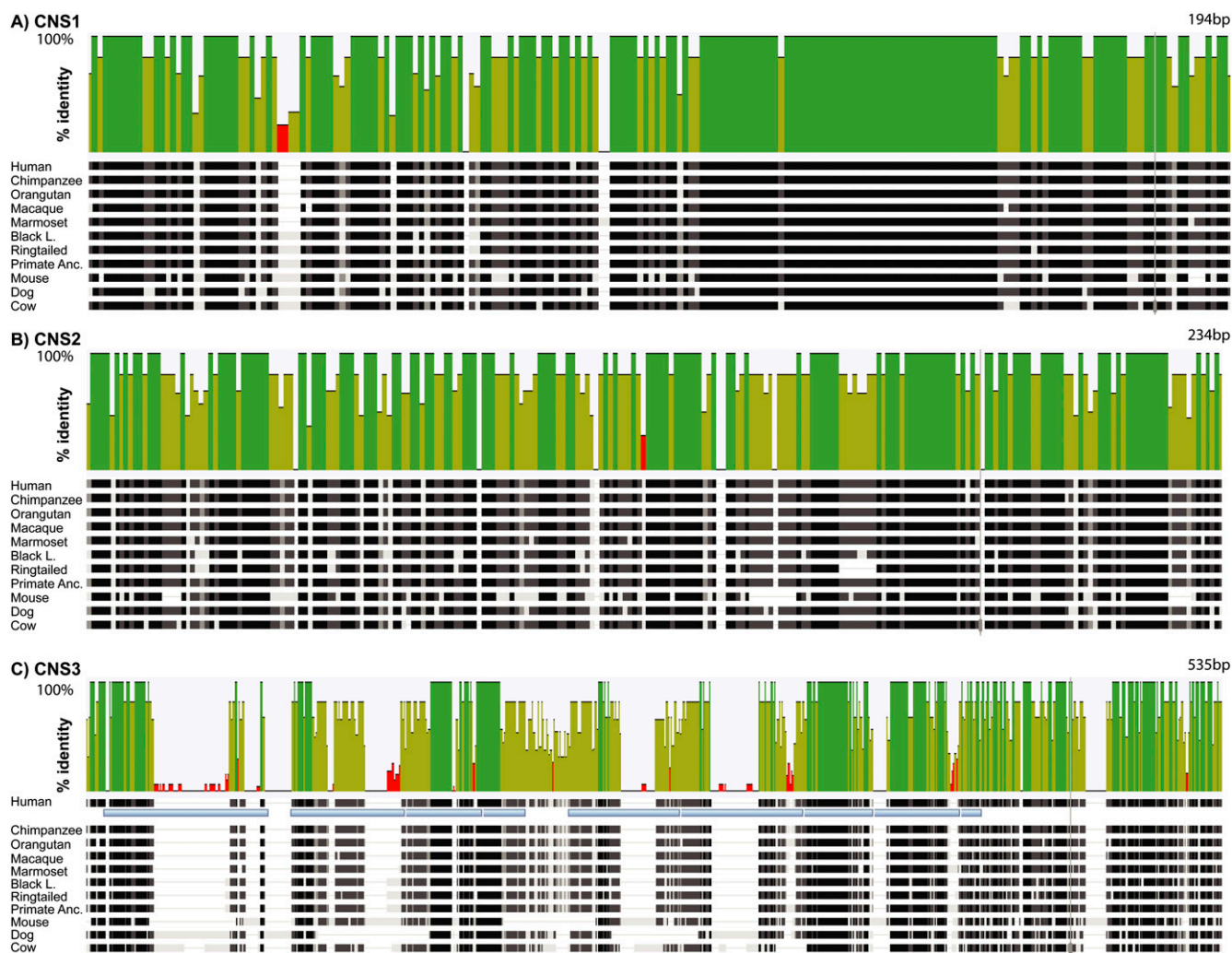


Figure 6. Multispecies alignments of conserved noncoding regions. Schematics of alignments across conserved noncoding sequences (CNS) are shown. See Figure 5 for the approximate locations within *XIST*. Along the top of each alignment is the sequence identity plot (from zero to 100% identity) comparing these sequences across all species and the reconstructed primate ancestor. (Green peaks) 100% identity among all species; (yellow and red peaks) lower identities. (A) CNS1 spans 194 bp and covers most of *XIST/Xist* exon 4. (B) CNS2 spans 243 bp. (C) CNS3 spans 535 bp and covers the *XIST/Xist* A repeat. (Light blue bars below the human sequence) The human A-repeat monomer units.

To conduct partial sequence comparisons, PCR products from the primer set spanning exons 1–6 from all lemurs were subcloned and sequenced. These sequences were aligned to those of human, mouse, and cow to compare the expressed regions (Fig. 7). Regions corresponding to human exons 1, 4, 5, and 6 were expressed in all species examined, while the other exons were expressed only in a subset of the female cell lines. Black lemur has the fewest expressed *XIST* exons in this analysis.

Comparative *TSIX* structure

While the critical role of *XIST/Xist* in X inactivation is clear, the involvement of other genes and loci is variable between species. For example, both human and mouse have *TSIX/Tsix* loci, but they have different gene structures (see Figs. 1, 3A) and seem to play very different roles in X inactivation (Migeon et al. 2002). In

mouse, *Tsix* plays a role in X-chromosome choice (Lee et al. 1999a), but there is no evidence to suggest that it is important in human X-chromosome choice. By extending *TSIX/Tsix* comparisons among many primates, it is clear that there are many repeat element insertions (Fig. 2B) in different lineages, and some species (e.g., ring-tailed lemur) have large deletions breaking up the overall structure. The 3' end of *TSIX*, which overlaps the 3' end of *XIST/Xist* on the opposite strand, is the most conserved in all species (Figs. 2B, 3B), while the middle region likely emerged within the past 90 Myr since it is only seen in the primates.

In mouse, the *DXPas34* locus (Debrand et al. 1999; Cohen et al. 2007) and the X-inactivation intergenic transcription element (*Xite*) have been reported to regulate *Tsix* gene expression. A previous study suggested that a portion of *DXPas34* (termed the "A region") is recognizable in humans (Cohen et al. 2007). This A region is orthologous in human and chimpanzee but does not

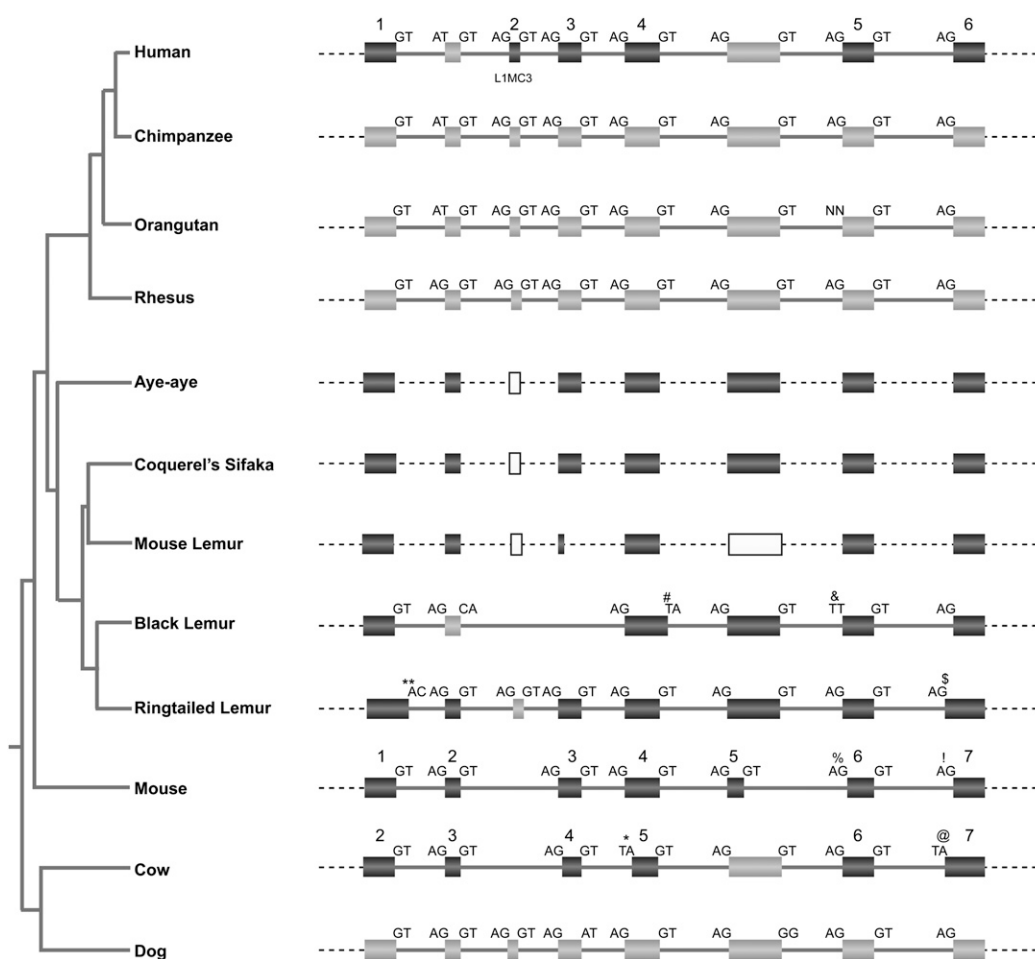


Figure 7. Comparative *XIST* expression. A schematic of *XIST* exons transcribed from human (based on NR_001564) and lemur (aye-aye, Coquerel's sifaka, gray mouse lemur, black lemur, and ring-tailed lemur) fibroblast cell lines is shown relative to known mouse and cow exons (NR_001463 and NR_001464, respectively) (Chureau et al. 2002). Numbers above exons correspond to known numbered exons in human, mouse, and cow based on these RefSeqs. The general phylogeny along the left was compiled from Murphy et al. (2007), Horvath et al. (2008), and Perelman et al. (2011). (Dark gray boxes) Exons transcribed in a species; (light gray boxes) conservation of DNA sequence with no verified expression. Putative splice donor and acceptor sites are indicated for each exon–intron splice junction. (Open boxes) Regions that are not expressed in the lemurs and for which there is not a complete genome sequence. (NN) The orangutan genome sequence has a short gap of N's at the putative start of this exon. (#) The black lemur exon ends 5 bp downstream from all other species (except mouse, which extends 4 bp downstream) and the following splice junction is a TA instead of a GT. (&) The black lemur exon starts 3 bp upstream and the adjacent splice junction is TT instead of AG. (\$) The ring-tailed lemur starts 96 bp upstream of human and black lemur. (**) The ring-tailed lemur exon extends 8 bp past all others. (%) The mouse exon starts 16 bp downstream from the human, cow, and ring-tailed lemur orthologous exon. (!) The mouse exon starts 271 bp upstream of the human and the black lemur exons. (*) The cow exon starts 67 bp downstream with a TA instead of AG at the splice junction. (@) The cow exon starts 64 bp upstream of the human and black lemur.

align in any other species in this analysis (Supplemental Fig. 3). Similarly, the mouse *DXPas34* region (including the A1, A2, and B regions) (Cohen et al. 2007) is not conserved at the sequence level in any other species in this analysis (Supplemental Fig. 3). Sequence across the mouse *Xite* region does not align well in any other species, suggesting that if any primates have a functional *Xite* locus it is not recognizable by sequence alignment. A lack of sequence identity does not necessarily indicate a lack of function, and these human–mouse variations further reinforce some of the distinct differences between rodent and primate X inactivation.

Marks of an inactive X chromosome in lemurs

XIST expression specific to female lemur cell lines is consistent with the expectation that female lemurs have an inactive X chromosome. To verify this cytologically, male and female fibroblast cell lines were assessed by immunofluorescence assays, using an antibody specific for the dimethylated form of histone H3, H3K4me2, which is deficient on the inactive X chromosome in humans (Boggs et al. 2002; Chadwick and Willard 2002) and mice (Heard et al. 2001). In male cell lines from both lemur species, all chromosomes appear consistently stained with the antibody to H3K4me2 (Supplemental Fig. 9c,i). In contrast, one X chromosome in the corresponding female cell lines is virtually devoid of antibody staining (Supplemental Fig. 9f,l), consistent with the presence of an inactive X chromosome in these cell lines.

Discussion

Comparative analyses of *XIST*, *TSIX*, and the candidate XIC offer insight into recent structural changes and evolutionarily conserved regions among primate genomes. While previous comparisons between mouse and human *XIST/Xist* and XIC/Xic have shed light on mechanisms of X inactivation, significant differences between human and mouse have prevented direct inferences between genomic sequence and function. Our approach using diverse primate comparisons has been informative in several areas.

Confidence in ancestral reconstructions

For XIC ancestral reconstructions, confidence in the presence or absence of a nucleotide is much higher than confidence in which base pair existed in an ancestor (Supplemental Fig. 4b). This is not surprising given that these are reconstructions across diverse mammalian species and that parts of the alignment are problematic due to low conservation and/or alignment difficulties (Chen and Tompa 2010). Therefore, for XIC reconstructions, we focused on the presence or absence of a base pair and did not focus on the exact nucleotide at each position in the ancestral sequence. Reconstructions across *XIST* show much higher confidence than those across the entire XIC region (cf. Supplemental Figs. 4 and 5), with regions of higher conservation in *XIST* seen across all species. The region with lowest confidence (position 24,000–26,000) spans part of the D repeat, which is known to vary substantially between species. Some species (e.g., the black lemur) entirely lack this repeat, while other species (e.g., cow) have a D-repeat region that does not align well to the other species, making reconstruction challenging (data not shown).

Recent gene and landscape restructuring in the XIC region

The candidate primate XIC has clearly undergone expansions and contractions along different lineages since the two lemurs have

a much more compact sequence across the XIC region than the rest of the primates (Figs. 2A, 3A; Supplemental Fig. 3). These landscape changes in the close vicinity of *XIST/Xist* further reinforce the malleability of this region, as has been underscored previously by comparisons of the mouse, marsupial, and monotreme genomes (Duret et al. 2006; Hore et al. 2007b). Interspersed repeats have played a role in differentially shaping primate genomes (Liu 2003; Liu et al. 2009), and analysis of total repeat content across this region is in agreement with this conclusion (Table 1). Comparisons of specific classes of repeats in the lemurs and the inferred ancestral primate XIC region indicate that the lemurs and ancestral primate have a lower *Alu*, MIR, and LINE content relative to the rest of the primates (Table 1), while some of the other repeats tend to fluctuate with lineage-specific trends. Since LINE elements have been proposed as way stations for *XIST* RNA (Lyon 1998; Bailey et al. 2000; Chow et al. 2010), it will be interesting to see if the lower LINE content among lemurs has any impact on X inactivation in these species.

Comparative structure of *XIST*

Reconstruction of the ancestral primate *XIST* gene highlights that the overall structure and content of the *XIST* locus have not significantly changed throughout primate evolution, but that the underlying sequence has been under low sequence constraint, as previously proposed (Hendrich et al. 1997; Nesterova and Slobodyanyuk 2001; Chureau et al. 2002). This further highlights the benefit of diverse primate comparisons, since, although the evolutionary distance between human and mouse (and dog and cow) is not that much greater than between human and lemur (>90 Mya vs. >80 Mya) (Murphy et al. 2007; Horvath et al. 2008; Perelman et al. 2011), the mouse region has undergone many more structural and repeat changes.

Since previous analyses of human, mouse, and cow highlighted species-specific exons (Chureau et al. 2002), it was informative to see which exons were conserved among the lemurs (Fig. 7). Given the high levels of *XIST/Xist* alternative splicing previously noted (Brown et al. 1991a, 1992), it was not surprising that not all exons were identified in all species using an expression-based approach. It is important to note that any single lemur species would not represent the level of diversity obtained by comparing all lemurs. This is even apparent when comparing two lemurs from the same taxonomic family (black lemur and ring-tailed lemur). It is informative that the black lemur has the fewest expressed exons and is also the only species so far identified that does not have the *XIST* D repeat; this suggests that the apparently missing exons and the D repeat are not critical for the process of X inactivation in black lemur (and potentially, therefore, in other species). These genomic and cDNA comparisons indicate that the second and third exons, which are missing from the black lemur transcript in Figure 7, are also absent entirely from the black lemur genome.

XIC noncoding RNAs

First described more than 20 yr ago (Brannan et al. 1990; Brown et al. 1991a), noncoding RNAs are now known to be prevalent around the genome and have been suggested to play functional roles in a variety of genomic, epigenetic, and developmental processes, with different evolutionary forces acting on them (Pang et al. 2006; Caley et al. 2010). Our analysis identified regions of orthology within *XIST/Xist*, *TSIX/Tsix*, and *NCRNA00183/JPX/Jpx/Enox* (recently shown to be an *Xist* activator in mouse) (Tian et al. 2010)

among all species (Fig. 3B). In contrast, we identified little or no orthology between mouse and primate sequences across other loci such as *Xite* and *DXPas34*, and only minimal orthology across one exon of *Tsx* between mouse and the lemurs. The low level or lack of sequence identity across these loci does not necessarily indicate a lack of function, as it is a general feature of many noncoding RNAs (Pang et al. 2006; Caley et al. 2010). One explanation for this trend is that noncoding RNAs may interact through higher-order structures and not directly through the underlying sequence (Caley et al. 2010). This, therefore, may be another level of species specificity and functional difference that will be important to characterize in a wider set of species and as more noncoding RNAs are characterized within the XIC (Tian et al. 2010).

Lemurs as potential models for X-inactivation studies

Our comparative analyses with lemurs suggest that they may be informative models for the further study of X inactivation. Significant differences in the candidate XIC region and in both *TSIX* and *XIST* structure suggest the possibility that X-inactivation mechanisms and/or the extent of transcriptional silencing in lemurs and other primates might also be different. Some lemur species can interbreed and form viable hybrid offspring (Horvath and Willard 2007). These female hybrid offspring would have X chromosomes from two parental species that would have many more sequence changes than any two individuals of the same species; much as interspecific mouse crosses have been valuable for the study of murine X inactivation (Yang et al. 2010), such sequence differences could be exploited to infer the silence or escape status of each gene in lemurs, as has been done on a comprehensive scale for human and mouse (Carrel and Willard 2005; Yang et al. 2010).

Methods

Lemur cell lines and DNA sampling

Lemur cell lines were obtained through Coriell Cell Repositories (<http://ccr.coriell.org/>) and the Integrated Primate Biomaterials and Information Resource (IPBIR) Collection (Supplemental Table 3). For all species, a male and female pair was available, with the exception that only a female gray mouse lemur cell line was available. Black lemur blood and buccal cells for DNA extraction were obtained from the Duke Lemur Center under research project BS-4-06-1 and Institutional Animal Care and Use Committee (IACUC) project A094-06-03.

DNA and RNA isolation

Bacterial artificial chromosomes (BACs) were obtained through BACPAC Resources (<http://bacpac.chori.org/>) as bacterial stabs. Single colonies were streaked onto LB with 12.5 μ g/mL chloramphenicol plates, and single colonies were used to inoculate LB media. BAC miniprep DNA was isolated with the Perfectprep BAC 96 kit (Eppendorf) and resuspended in water according to the manufacturer. Genomic DNA was isolated from cell lines using the Genra PUREGENE kit according to the manufacturer's recommendations. RNA was isolated from cell lines using the Genra VERSAGENE kit according to the manufacturer's recommendations.

Lemur BAC library hybridization

Four primer pairs (probes A–D, as shown in Supplemental Fig. 1) were designed to regions conserved between human and dog that

were ~100 to 150 kb apart in the candidate XIC region in humans (Table 1). Primer pairs were used in PCR assays with black lemur (*Eulemur macaco macaco*) and ring-tailed lemur (*Lemur catta*) genomic DNA from IPBIR cell line PRO0254 and Coriell cell line AG07100, respectively. PCR products were purified with the Roche Diagnostics Corporation High Pure PCR Product Purification Kit, and 25–50 ng of purified DNA was individually labeled with [α - 32 P]dCTP using the High Prime DNA labeling Kit (Roche Diagnostics Corporation). BAC library membranes from the black and ring-tailed lemurs (CHORI-273 and LBNL-2, respectively [BACPAC Resources]) were hybridized as described previously (Horvath et al. 2003). Hybridized membranes were imaged for at least 16 h using a PhosphorImager (Amersham Biosciences), and positives were called by hand.

Lemur sequence across the primate candidate XIC

Lemur BAC sequences were generated by the NIH Intramural Sequencing Center (NISC). The BACs were shotgun sequenced and assembled to “comparative-grade” standards, with all contigs ordered and oriented (Blakesley et al. 2004). Black lemur sequence across the candidate primate XIC region was obtained by concatenating accession number AC204188.2 (CHORI-273 BAC 137L18) from base pairs 1 to 138,202 with accession number AC203493.2 (CHORI-273 BAC 16H20) from base pair 1 to the end of the clone, for a total of 335,784 bp. Similarly, the orthologous sequence of the ring-tailed lemur was obtained by concatenation of accession number AC204810.2 (LBNL-2 clone 143K1) from position 1 to 128,783 and accession number AC203729.2 (LBNL-2 clone 223D18) from position 1 to the end of the clone for a total of 288,342 bp. The sequence gaps at position 110359–110458 within AC203493.2 (FJ156096) and the gaps at position 99465–99564 (FJ156094) and position 101792–101891 (FJ156095) within AC203729.2 were filled by long-range PCR and sequence analysis of at least three subclones across the gap region. See Supplemental Table 4 for a list of primer names and sequences.

Computational analyses

RepeatMasked (A Smit and P Green, RepeatMasker version 07/13/2002; <http://www.repeatmasker.org>) BACs were aligned using mVISTA (Mayor et al. 2000; Brudno et al. 2003; Frazer et al. 2004) and Geneious V4.8.5 (Drummond et al. 2009). All non-lemur mammalian sequences were downloaded from the UCSC Genome Browser (Kent et al. 2002) for both the candidate XIC region and the *XIST/Xist* locus. For this analysis, coordinates for the candidate XIC were from hg19_chrX:72661881–73165617, panTro2_chrX:72775307–73287735, ponAbe2_chrX:70855024–71384876, rheMac2_chrX:72443776–73067869, calJac3 concatenated from chrX_GL286110_random:10704–43855, chrX:65103379–65242357, chrX_GL286112_random:82674–99685, chrX:65242358–65554460, black lemur from 1 to 311,784 bp in the above concatenated sequence, ring-tailed lemur from 1 to 264,844 bp in the above concatenated sequence, mm9_chrX:100506112–100690172, canFam2_chrX:60167770–60454393, and bosTau4_chrX:47140322–47488480 (reverse complemented). Coordinates extracted for the *XIST/Xist* region were as follows: hg19_chrX:73040491–73072588, panTro2_chrX:73151360–73182518, ponAbe2_chrX:71263811–71295067, rheMac2_chrX:72942857–72975346, calJac3_chrX:65378813–65411782, black lemur: coordinates 226839–54509 extracted from the above concatenated sequence, ring-tailed lemur: coordinates 208158–238305 extracted from the above concatenated sequence, mm9_chrX:100655710–100678598, canFam2_chrX:60374075–60411096, and bosTau4_chrX:47179805–47216560 (reverse complemented). Sequences were globally aligned using

MLAGAN (Brudno et al. 2003) with the evolutionary tree [((((((Hum, Chimp) Orang) Rhesus) Marmoset)(BlkLem, Ringtail) Mouse)(Dog, Cow))] constructed using a maximum likelihood approach with branch lengths estimated in PAUP* 4.0a109 (Swofford 2002). The accession numbers used for annotations were as follows: human [*CDX4* (NM_005193), *CHIC1* (NM_001039840), *DXPas34* A region (Cohen et al. 2007), *TSIX* (NR_003255), *XIST* (NR_001564), *JPX/ENOX* (NR_024582)], mouse [*Cdx4* (NM_007674), *Chic1* (NM_009767), *Tsx* (NM_009440), *DXPas34* region (mm9_chrX:100643413–100644865 coordinates as determined by Dotter plots) (Chao et al. 2002; Cohen et al. 2007; data not shown), *Xite* region (genomic region in AY190761), *Tsix* (NR_002844), *Xist* (NR_001463), *Jpx* (exon 1 from AK148110 represents exon 1 here and exon 1 from AK050201 represents exon 2 here)], and cow [*Xist* (NR_001464)]. RankVISTA output for the three conserved non-coding regions was as follows: CNS 1, $p = 3.9 \times 10^{10}$; CNS2, $p = 1.3 \times 10^6$; CNS3, $p = 2 \times 10^8$.

Ancestral reconstructions were carried out in two steps using all the above sequences. First, the presence or absence of a nucleotide at each position for the different ancestors was computed using the phylo-HMM approach described in Diallo et al. (2007). This method allows the computation of the posterior distribution of insertion and deletion scenarios. Second, the nucleotide annotation was performed using a standard continuous time DNA nucleotide model as described in Blanchette et al. (2008). Similar to several other studies on mammalian sequences, the HKY model of evolution was chosen (Blackburn 1991; Kim and Sinha 2007; Paten et al. 2008). Alignments were analyzed for conserved regions and annotated using Geneious V4.8.5 (Drummond et al. 2009). *XIST* repeat regions were identified using the Yen et al. (2007) coordinates and Tandem Repeats Finder v4.03 (Benson 1999).

The confidence values of ancestral reconstructions were computed in two different steps. Confidence in the presence or absence of a nucleotide at each position in the different ancestors was computed using the forward-backward algorithm within the phylogenetic-HMM (see Diallo et al. 2007). The values are indicated in terms of probabilities between 1 and 100. The second step computed the confidence level of substitution using a variant of the Felsenstein algorithm (Blanchette et al. 2008). It is a ratio between the predicted likelihood of the predicted ancestral nucleotide and the sum of the overall likelihood of all possible alternative scenarios.

Synthesis of cDNA

Approximately 1 μ g of total RNA was treated with 1 unit of RNase-free DNase I (New England Biolabs) for 10 min at 37°C. EDTA was added to a final concentration of 0.4 mM, and heat inactivation of the DNase I proceeded for 10 min at 75°C. First-strand cDNA synthesis using the entire DNase-treated RNA volume proceeded with the addition of the following reagents: DTT (Invitrogen) to a final concentration of 0.01 M, 1 mM each dNTP (Invitrogen), 0.5 μ L of Random Hexamer (Amersham Biosciences), 20 units of RNase OUT (Invitrogen), and 200 units of MMLV Reverse Transcriptase (Invitrogen) was conducted for 10 min at 25°C followed by incubation for 2 h at 42°C and final heat inactivation of the enzymes for 10 min at 95°C. A no-RT control with the addition of water instead of MMLV-RT was conducted for each cDNA synthesis reaction. Subsequent PCR assays contained 1 μ L (1/20 volume) of first-strand synthesized cDNA or water as a negative control.

PCR and sequencing

BAC-end sequencing reactions were conducted using 10 μ L of Perfectprep BAC 96 (Eppendorf) template heat-denatured for 5 min at 95°C followed by the immediate addition on ice of 3 μ L

of BigDye v3.1 (Applied Biosystems), 3 μ L of 5 \times reaction buffer, 0.5 μ M primer, 0.75 mM MgCl₂, and water to a final volume of 20 μ L. Cycle sequencing was performed using 100 cycles of 95°C for 15 sec, 50°C for 15 sec, followed by 60°C for 4 min and a final hold at 4°C. Primers EPT7 and EPSP6 were used for BAC end sequencing. The quality of sequence data was assessed using PHRED/PHRAP/CONSED software (<http://genome.wustl.edu>). PCR and sequencing reactions were carried out as previously described (Horvath et al. 2008) using 0.625 U of Taq Polymerase (Invitrogen) with a 72°C extension. Long-range PCR was conducted for primer pairs JHX87/JHX94, JH464/JH465, and JH466/JH468 using the Roche Diagnostics Corporation Expand Long Template PCR System according to the manufacturer's recommendations. PCR products from primer pairs JHX47/JHX52, JHX87/JHX94, JH464/JH465, and JH466/JH468 (Supplemental Table 3) were TA-cloned using the pGEM-T Easy Vector System II (Promega) and sequenced from the plasmid as previously described (Horvath et al. 2008). Accession numbers corresponding to these long-range PCR products (FJ156094-96) were deposited in GenBank.

Immunostaining

Metaphase spreads were obtained from exponentially growing cells after 1 to 2 h of colcemid treatment using standard protocols. Slides were fixed in a 4% formaldehyde–1 \times PBS–0.1% Triton solution for 10 min. The slides were then washed twice in 1 \times PBS for 2 min before the addition of antibodies.

To detect epigenetic marks characteristic of an inactive X chromosome, we used a 1:200 dilution of the primary antibody, rabbit monoclonal Anti-H3K4me2 (Upstate Cat. No. 05-790), and a 1:200 dilution of the secondary antibody, Cy3-conjugated donkey anti-rabbit IgG (Jackson Laboratories Cat. No. 711-165-152). Immunostaining was carried out using minor modifications to procedures described previously (Chadwick and Willard 2001).

Fluorescence in situ hybridization

Isolation of BAC DNA was performed using a QIAGEN Maxiprep Kit. One microgram of BAC DNA was labeled with Spectrum Green dUTP (Abbott Molecular) using the Nick Translation Reagent Kit (Abbott Molecular). Probes were precipitated with the addition of 10 μ g of CotI DNA and rehydrated in 10 μ L of Hybrisol VII (MP Biomedicals) for 2–16 h at 37°C. Probes were denatured for 7 min at 72°C and then placed for 30–90 min at 37°C. For each cell line, we scored 12–30 metaphase spreads for staining of X chromosomes with an antibody raised to H3K4me2.

Slides previously immunostained were washed once in 1 \times PBS–0.05% Tween for 5 min and then denatured one at a time in 70% formamide–2 \times SSC (pH 7.0) for 12 min at 75°C. Slides were briefly washed in 1 \times PBS–0.05% Tween, and 14–20 μ L of denatured probe was added per slide. Slides were then placed in a humid chamber and hybridized overnight at 37°C. Post-hybridization washes consisted of two 8-min washes in 50% formamide–2 \times SSC (pH 7.0) at 42°C, then one 8-min wash in 2 \times SSC at 37°C. Slides were briefly rinsed in reagent-grade water before being counterstained with 4,6-diaminidino-2-phenylindole in Vectashield (Vector Laboratories). Slides were analyzed under a Zeiss Axiovert 200M microscope fitted with a Hamamatsu ORCA-ER camera. Images were captured with OpenLab (Improvision) and processed with Adobe Photoshop.

Acknowledgments

We thank Beth Sullivan and members of the Willard laboratory for technical discussions. J.E.H. thanks David Haring of the Duke

Lemur Center for providing lemur photographs. This is Duke Lemur Center Publication Number 1199. The work was supported in part by the Intramural Research Program of the National Human Genome Research Institute, National Institutes of Health, and by funds from the Genome Biology Program of the Duke Institute for Genome Sciences & Policy (IGSP). J.E.H. was supported in part by a fellowship from the IGSP Center for Evolutionary Genomics.

Authors' contributions: J.E.H. and H.F.W. designed the study. J.E.H. performed experiments, analyzed results, and wrote the paper. C.B.S. performed data analysis and assisted with figure preparation and manuscript revisions. S.L.M. performed IF FISH experiments and PCR analysis. D.L.S. conducted phylogenetic comparisons. A.B.D. conducted ancestral reconstructions. The NISC and E.D.G. generated the BAC sequences. H.F.W. and E.D.G. contributed to data interpretation, helped with manuscript preparation, and provided financial support.

References

- Bailey JA, Carrel L, Chakravarti A, Eichler EE. 2000. Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: The Lyon repeat hypothesis. *Proc Natl Acad Sci* **97**: 6634–6639.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**: 573–580.
- Blackburn E. 1991. Structure and function of telomeres. *Nature* **350**: 569–573.
- Blakesley RW, Hansen NF, Mullikin JC, Thomas PJ, McDowell JC, Maskeri B, Young AC, Benjamin B, Brooks SY, Coleman BI, et al. 2004. An intermediate grade of finished genomic sequence suitable for comparative analyses. *Genome Res* **14**: 2235–2244.
- Blanchette M, Green ED, Miller W, Haussler D. 2004. Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res* **14**: 2412–2423.
- Blanchette M, Diallo AB, Green ED, Miller W, Haussler D. 2008. Computational reconstruction of ancestral DNA sequences. *Methods Mol Biol* **422**: 171–184.
- Boffelli D, McAuliffe J, Ovcharenko D, Lewis KD, Ovcharenko I, Pachter L, Rubin EM. 2003. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. *Science* **299**: 1391–1394.
- Boggs B, Cheung P, Heard E, Spector D, Chinnault AC, Allis CD. 2002. Differentially methylated forms of histone H3 show unique association patterns with inactive human X chromosomes. *Nat Genet* **30**: 73–76.
- Borsani G, Tonlorenzi R, Simmler MC, Dandolo L, Arnaud D, Capra V, Grompe M, Pizzuti A, Muzny D, Lawrence C, et al. 1991. Characterization of a murine gene expressed from the inactive X chromosome. *Nature* **351**: 325–329.
- Brannan CI, Dees EC, Ingram RS, Tlghman SM. 1990. The product of the H19 gene may function as an RNA. *Mol Cell Biol* **10**: 28–36.
- Brockdorff N. 2002. X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* **18**: 352–358.
- Brockdorff N, Ashworth A, Kay GE, Cooper P, Smith S, McCabe VM, Norris DP, Penny GD, Patel D, Rastan S. 1991. Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**: 329–331.
- Brockdorff N, Ashworth A, Kay GE, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. 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.
- Brown CJ, Baldry SE. 1996. Evidence that heteronuclear proteins interact with XIST RNA in vitro. *Somat Cell Mol Genet* **22**: 403–417.
- Brown CJ, Ballabio A, Rupert JL, Lafreniere RG, Grompe M, Tonlorenzi R, Willard HF. 1991a. A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**: 38–44.
- Brown CJ, Lafreniere RG, Powers VE, Sebastio G, Ballabio A, Pettigrew AL, Ledbetter DH, Levy E, Craig IW, Willard HF. 1991b. Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**: 82–84.
- Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. 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.
- Bruđno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* **13**: 721–731.
- Caley DP, Pink RC, Trujillano D, Carter DR. 2010. Long noncoding RNAs, chromatin, and development. *ScientificWorldJournal* **10**: 90–102.
- Carrel L, Willard HF. 2005. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**: 400–404.
- Chadwick BP, Willard HF. 2001. A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol* **152**: 375–384.
- Chadwick B, Willard HF. 2002. Cell cycle-dependent localization of macroH2A in chromatin of the inactive X chromosome. *J Cell Biol* **157**: 1113–1123.
- Chadwick B, Willard HF. 2003. Barring gene expression after XIST: maintaining facultative heterochromatin on the inactive X. *Semin Cell Dev Biol* **14**: 359–367.
- Chao W, Huynh KD, Spencer RJ, Davidow LS, Lee JT. 2002. CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science* **295**: 345–347.
- Chen X, Tompa M. 2010. Comparative assessment of methods for aligning multiple genome sequences. *Nat Biotechnol* **28**: 567–572.
- Cheng Z, Ventura M, She X, Khaïtovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, et al. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature* **437**: 88–93.
- Chow JC, Ciaudo C, Fazzari MJ, Mise N, Servant N, Glass JL, Attreed M, Avner P, Wutz A, Barillot E, et al. 2010. LINE-1 activity in facultative heterochromatin formation during X chromosome inactivation. *Cell* **141**: 956–969.
- Chureau C, Prissette M, Bourdet A, Barbe V, Cattolico L, Jones L, Eggen A, Avner P, Duret L. 2002. Comparative sequence analysis of the X-inactivation center region in mouse, human, and bovine. *Genome Res* **12**: 894–908.
- Cohen D, Davidow L, Erwin J, Xu N, Warshawsky D, Lee J. 2007. The DXPas34 repeat regulates random and imprinted X inactivation. *Dev Cell* **12**: 57–71.
- Debrand E, Chureau C, Arnaud D, Avner P, Heard E. 1999. Functional analysis of the DXPas34 locus, a 3' regulator of Xist expression. *Mol Cell Biol* **19**: 8513.
- Diallo AB, Makarenkov V, Blanchette M. 2007. Exact and heuristic algorithms for the Indel Maximum Likelihood Problem. *J Comput Biol* **14**: 446–461.
- Diallo AB, Makarenkov V, Blanchette M. 2010. Ancestors 1.0: a web server for ancestral sequence reconstruction. *Bioinformatics* **26**: 130–131.
- Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2009. *Geneious v4.7*. <http://www.geneious.com>.
- Duret L, Chureau C, Samain S, Weissenbach J, Avner P. 2006. The Xist RNA gene evolved in eutherians by pseudogenization of a protein-coding gene. *Science* **312**: 1653–1655.
- Elisaphenko E, Kolesnikov N, Shevchenko A, Rogozin I, Nesterova T, Brockdorff N, Zakian S, Gadagkar S. 2008. A Dual origin of the Xist gene from a protein-coding gene and a set of transposable elements. *PLoS ONE* **3**: e2521. doi: 10.1371/journal.pone.0002521.
- Erwin J, Lee J. 2008. New twists in X-chromosome inactivation. *Curr Opin Cell Biol* **20**: 349–355.
- Ferguson-Smith M, Trifonov V. 2007. Mammalian karyotype evolution. *Nat Rev Genet* **8**: 950–962.
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. *Nucleic Acids Res* **32**: W273–W279.
- Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP. 1998. Toward a phylogenetic classification of primates based on DNA evidence complemented by fossil evidence. *Mol Phylogenet Evol* **9**: 585–598.
- Heard E, Kress C, Mongelard F, Courtier B, Rougeulle C, Ashworth A, Vourc'h C, Babinet C, Avner P. 1996. Transgenic mice carrying an Xist-containing YAC. *Hum Mol Genet* **5**: 441–440.
- Heard E, Mongelard F, Arnaud D, Avner P. 1999. Xist yeast artificial chromosome transgenes function as X-inactivation centers only in multicopy arrays and not as single copies. *Mol Cell Biol* **19**: 3156–3166.
- Heard E, Rougeulle C, Arnaud D, Avner P, Allis C, Spector D. 2001. Methylation of histone H3 at Lys-9 is an early mark on the X chromosome during X inactivation. *Cell* **107**: 727–738.
- Hendrich BD, Plenge RM, Willard HF. 1997. Identification and characterization of the human XIST gene promoter: implications for models of X chromosome inactivation. *Nucleic Acids Res* **25**: 2661–2671.
- Herzig LB, Romer JT, Horn JM, Ashworth A. 1997. Xist has properties of the X-chromosome inactivation centre. *Nature* **386**: 272–275.
- Hore T, Rapkins R, Graves J. 2007a. Construction and evolution of imprinted loci in mammals. *Trends Genet* **23**: 440–448.
- Hore TA, Koina E, Wakefield M, Marshall Graves J. 2007b. The region homologous to the X-chromosome inactivation centre has been

- disrupted in marsupial and monotreme mammals. *Chromosome Res* **15**: 147–161.
- Horvath JE, Willard HF. 2007. Primate comparative genomics: lemur biology and evolution. *Trends Genet* **23**: 173–182.
- Horvath JE, Gulden CL, Bailey JA, Yohn C, McPherson JD, Prescott A, Roe BA, De Jong PJ, Ventura M, Misceo D, et al. 2003. Using a pericentromeric interspersed repeat to recapitulate the phylogeny and expansion of human centromeric segmental duplications. *Mol Biol Evol* **20**: 1463–1479.
- Horvath J, Weisrock D, Embry S, Fiorentino I, Balhoff J, Kappeler P, Wray G, Willard H, Yoder A. 2008. Development and application of a phylogenomic toolkit: Resolving the evolutionary history of Madagascar's lemurs. *Genome Res* **18**: 489–499.
- Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996–1006.
- Kim J, Sinha S. 2007. Indelign: a probabilistic framework for annotation of insertions and deletions in a multiple alignment. *Bioinformatics* **23**: 289–297.
- Lafreniere RG, Brown CJ, Rider S, Chelly J, Taillon-Miller P, Chinault AC, Monaco AP, Willard HF. 1993. 2.6 Mb YAC contig of the human X inactivation center region in Xq13: physical linkage of the RPS4X, PHKA1, XIST and DXS128E genes. *Hum Mol Genet* **2**: 1105–1115.
- Lee JT, Strauss WM, Dausman JA, Jaenisch R. 1996. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* **86**: 83–94.
- Lee JT, Davidow LS, Warshawsky D. 1999a. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* **21**: 400–404.
- Lee JT, Lu N, Han Y. 1999b. Genetic analysis of the mouse X inactivation center defines an 80-kb multifunction domain. *Proc Natl Acad Sci* **96**: 3836–3841.
- Leppig KA, Brown CJ, Bressler SL, Gustashaw K, Pagon RA, Willard HF, Disteche CM. 1993. Mapping of the distal boundary of the X-inactivation center in a rearranged X chromosome from a female expressing XIST. *Hum Mol Genet* **2**: 883–887.
- Liu G. 2003. Analysis of primate genomic variation reveals a repeat-driven expansion of the human genome. *Genome Res* **13**: 358–368.
- Liu GE, Alkan C, Jiang L, Zhao S, Eichler EE. 2009. Comparative analysis of *Alu* repeats in primate genomes. *Genome Res* **19**: 876–885.
- Lyon MF. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**: 372–373.
- Lyon M. 1998. X-chromosome inactivation: a repeat hypothesis. *Cytogenet Cell Genet* **80**: 133–137.
- Maenner S, Blaud M, Fouillen L, Savoye A, Marchand V, Dubois A, Sanglier-Cianferani S, Van Dorsseleer A, Clerc P, Avner P, et al. 2010. 2-D structure of the A region of Xist RNA and its implication for PRC2 association. *PLoS Biol* **8**: e1000276. doi: 10.1371/journal.pbio.1000276.
- Matsuura S, Episkopou V, Hamvas R, Brown SD. 1996. Xist expression from an Xist YAC transgene carried on the mouse Y chromosome. *Hum Mol Genet* **5**: 451–459.
- Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I. 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**: 1046–1047.
- Migeon B, Lee C, Chowdhury A, Carpenter H. 2002. Species differences in TSIX/Tsix reveal the roles of these genes in X-chromosome inactivation. *Am J Hum Genet* **71**: 286–293.
- Murphy WJ, Larkin DM, Everts-van der Wind A, Bourque G, Tesler G, Avuil L, Beever JE, Chowdhury BP, Galibert F, Gatzke L, et al. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**: 613–617.
- Murphy W, Pringle T, Crider T, Springer M, Miller W. 2007. Using genomic data to unravel the root of the placental mammal phylogeny. *Genome Res* **17**: 413–421.
- Nesterova T, Slobodyanyuk Y. 2001. Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res* **11**: 833–849.
- Okamoto I, Heard E. 2009. Lessons from comparative analysis of X-chromosome inactivation in mammals. *Chromosome Res* **17**: 659–669.
- Pang KC, Frith MC, Mattick JS. 2006. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet* **22**: 1–5.
- Paten B, Herrero J, Fitzgerald S, Beal K, Flicek P, Holmes I, Birney E. 2008. Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res* **18**: 1829–1843.
- Penny GD, Kay GF, Sheardown SA, Rastan S, Brockdorff N. 1996. Requirement for Xist in X chromosome inactivation. *Nature* **379**: 131–137.
- Perelman P, Johnson W, Roos C, Seuanez HN, Horvath JE, Moreira MAM, Kessing B, Pontius J, Roelke M, Rumpler Y, et al. 2011. A molecular phylogeny of living primates. *PLoS Genet* **7**: e1001342. doi: 10.1371/journal.pgen.1001342.
- Rastan S. 1983. Non-random X-chromosome inactivation in mouse X-autosome translocation embryos—location of the inactivation centre. *J Embryol Exp Morphol* **78**: 1–22.
- Swofford DL. 2002. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)*. Version 4. Sinauer, Sunderland, MA.
- Therman E, Sarto GE, Patau K. 1974. Center for Barr body condensation on the proximal part of the human Xq: a hypothesis. *Chromosoma* **44**: 361–366.
- Tian D, Sun S, Lee JT. 2010. The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* **143**: 390–403.
- Veyrunes F, Waters PD, Miethke P, Rens W. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Res* **18**: 965–973.
- Woodburne MO, Rich TH, Springer MS. 2003. The evolution of tribospheny and the antiquity of mammalian clades. *Mol Phylogenet Evol* **28**: 360–385.
- Wutz A, Rasmussen T, Jaenisch R. 2002. Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat Genet* **30**: 167–174.
- Yang F, Babak T, Shendure J, Disteche CM. 2010. Global survey of escape from X inactivation by RNA-sequencing in mouse. *Genome Res* **20**: 614–622.
- Yen ZC, Meyer IM, Karalic S, Brown CJ. 2007. A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* **90**: 453–463.
- Zhao J, Sun BK, Erwin JA, Song JJ, Lee JT. 2008. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**: 750–756.

Received June 17, 2010; accepted in revised form March 18, 2011.



Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate *XIST* locus

Julie E. Horvath, Christina B. Sheedy, Stephanie L. Merrett, et al.

Genome Res. 2011 21: 850-862 originally published online April 25, 2011

Access the most recent version at doi:[10.1101/gr.111849.110](https://doi.org/10.1101/gr.111849.110)

Supplemental Material <http://genome.cshlp.org/content/suppl/2011/03/25/gr.111849.110.DC1.html>

References This article cites 80 articles, 36 of which can be accessed free at:
<http://genome.cshlp.org/content/21/6/850.full.html#ref-list-1>

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
