EVOLUTION OF THE STRUCTURE OF THE XIST LOCUS IN MAMMALS

Kolesnikov N.N.*, Elisafenko E.A., Zakian S.M.
Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia
*Corresponding author: e-mail: kolesnikov@bionet.nsc.ru

Key words: X chromosome inactivation, gene XIST, structural organization, evolution, mammalian XIST gene ancestor

SUMMARY

Motivation: The origin of sex chromosomes and the emergence of the mechanism of inactivation of the X chromosomes in the course of mammalian evolution remain some of the most intriguing phenomena. X inactivation is controlled by a special locus of the X chromosome: the inactivation center, in which the central functional role is assigned to the Xist gene. Despite progress, X inactivation is still in the spotlight and still promises surprises. The question of the origin of X inactivation is opened-ended: what molecular mechanisms made it happen? We are particularly interested in the question of the origin of the Xist gene. How and when did this gene emerge? On the basis of what sequences did they? And since this gene is in the class of large non-coding regulatory RNA genes, this question is more general. The aim of this work was to reconstruct the ancestral Xist gene of placental mammals, to reveal features of its structure and identify its evolutionary pathways.

Results: We have performed a comparative analysis of the structure and organization of this gene in a variety of evolutionarily close and distant mammalian species; revealed the mosaic evolution of the structure elements of the gene; identified common conserved and species-specific gene sequences; revealed an interspecific lability of the exon-intron structure; determined the structure of the gene for chimpanzee, dog and rat; reconstructed the gene, which was ancestral to Xist in placental mammals and about 100 Myr ago consisted of 10 exons and had most probably emerged de novo with the advent of the sex chromosomes.

INTRODUCTION

Inactivation of one of the two X chromosomes in female mammals is an impressive and large-scale example of epigenetic regulation, in the course of which most genes of one X chromosome in the females become switched off. It has been postulated that this phenomenon is associated with dose compensation in mammalian females (Lyon, 1961). Transcriptional inactivation of one of the X chromosomes in mammalian females is a complex process. This process involves multiple chromatin modifications, which lead to the formation of stable facultative heterochromatin, and that heterochromatin is preserved in the succession of cell divisions.

This process is controlled by the inactivation center denoted as XIC (X – inactivation center; XIC, man; Xic, mouse), which spans about 1000 kb in the human X chromosome and about 450 kb in the mouse X chromosome. As has been demonstrated in recent years, XIC/Xic is a complex genetic locus, which contains important functional elements. Of them, the key element is the XIST/Xist gene (X-inactive specific transcript gene; XIST,
man; Xist, mouse). Transcription of this gene leads to the formation of non-coding RNAs spanning the entire chromosome and trigger a cascade of epigenetic chromatin modifications, which results in the inactivation in cis of the X chromosome (Borsani et al., 1991; Brockdorff et al., 1991; Brown et al., 1991).

The XIST/Xist gene contains exons and introns and does not contain large ORFs. XIST/Xist RNA is subject to processing in the form of splicing into alternative transcripts and polyadenylation and stays in the nucleus. Regulation of this gene, in turn, is exerted by cis-elements and in mouse additionally by the antisense transcript of the Xist gene: the Tsix gene (Lee et al., 1999). Human TsIX is reduced and functionally inactive compared to the mouse homologue, which suggests that regulation of the expression of the key gene must be different in different species (Migeon et al., 2002).

Not only the XIST/Xist gene is involved in the process of inactivation, but also belongs to a so far small class of large non-coding regulatory RNA (ncRNA), the study of which has only recently been initiated and is believed to be promising (Furuno et al., 2006).

MATERIALS AND METHODS

The common approach used consisted in analyzing the genomic sequences in the corresponding databases of sequenced genomes. The computer analysis was performed using the most recent versions of program packages such as BLAST (Altschul et al., 1990, http://www.ncbi.nlm.nih.gov/), which searches for homologous sequences; TRF (Benson, 1999), for tandem repeats; IRF (Warburton et al., 2004), for inverted repeats; RepeatMasker, (http://www.genome.washington.edu/UWGC/analysisstools/repeatmask.htm) for mobile elements; Fasta (Pearson, Lipman, 1988) and CLUSTALX (Jeanmogin et al., 1988), which align two or more sequences; the genomic analysis of large loci was performed using software programs and data available at servers (http://genome.ucsc.edu/; http://www.ensembl.org/) and the PipMaker software program (http://bio.cse.psu.edu).

The comparative analysis of the gene structure was performed on 10 species of four orders of placental mammals: Rodentia, Cetartiodactyla, Carnivora, Primates. We had previously determined the gene structure in four related vole species (Nesterova et al., 2001), the gene orthologs in three species: chimpanzee, dog and rat, have now been identified by multiple and pairwise alignments with known human, mouse and bovine sequences (Chureau et al., 2002).

RESULTS AND DISCUSSION

Gene structure. Overall, the structure of the XIST/Xist gene was found to be similar in all the species studied. However, of the 10 exons revealed, only six occur in all the species as part of mature RNA, and the others have species specific preferences. In doing so, we omitted the Xist RNA isoforms that resulted from alternative splicing from consideration. Thus, in man and mouse, this gene consists of 8 exons, of which only 6 are common to both. Exons 2 and 5 in man and chimpanzee are not what they are in rodents, but they are preserved in intron sequences in the inactive state (Table 1). We denoted the exons that are inactivated in some species but preserved in introns as pseudoexons (pEx). Three such pseudoexons, which are active in some species and inactivated in others, have been identified (Table 1). On the other hand, exon 2, which can be found in man, chimpanzee and dog, has been lost to the other species. The most interesting about it is that exon 2 expressed in man represents a small fragment of the 3′-region of L1MC3 transposon (Fig. 1), which is present in both chimpanzee and dog as part of the gene; however, as far as dog is concerned, this exon is inactive.
Figure 1. Comparative analysis of the Xist gene locus in placental mammalian species, the ancestral Xist gene versus genes in Homo sapiens (H.s.), Canis familiaris (C.f.), Bos taurus (B.t.), Rattus norvegicus (R.n.), Mus musculus (M.m.) and Microtus rossiaemeridionalis (M.r.). Percent identity plot (PIP) of the ancestral Xist gene relative to various species is shown on the X axis, and the percentage of its identity (50–100 %) to various species is shown on the Y axis.

This is a striking example of how a transposon participates in the formation of the structure of a regulatory gene in man and chimpanzee: part of the transposon sequence is recruited as an exon.

Alongside the formation of pseudoexons, which arise due to mutation to intron-exon sites (pEx2h, man; pEx2p - chimpanzee), we could observe exons shortened at 3' ends due to the formation of new 3'-5' exon/intron junctions: Ex2r (rat), Ex7m (mouse), Ex7v (vole).

Thus, the XIST/Xist structure is not strictly conserved or stable from an evolutionary point of view: there is a lability of exon/intron transitions, which might be associated with species specific features of the gene functioning.

XIST/Xist gene ancestor. Pairwise and multiple alignment allowed us to reconstruct the consensus sequence of the ancestor gene, which existed in placental mammals about 100 Myr ago at the onset of species radiation, and to assess how far its structural elements have diverged from the consensus. The Xist ancestor gene consisted of 10 exons and its genomic locus was about 30 kb in size (Table 1, Fig. 1). As can be seen from the table, the overall length of the gene varies from 21 to 37 kb with a tendency towards a lesser size of the locus in Rodentia species (21–23 kb) and a greater size in the other species (32–37.5 kb). The shortening of the overall locus length in rodents is largely due to the shortening of intron lengths. The total length of exons is more than that of introns and the ratio of exon/intron length varies from 1.2 in man to 2.3 in cow, which makes a difference between the Xist gene and protein-encoding genes, in which the introns are ten to hundreds of times as long as the exons.
### Table 1. Comparative size (in base pairs) of the XIST/Xist locus and its structural elements

<table>
<thead>
<tr>
<th>Species</th>
<th>Homologous exons</th>
<th>Total intron length</th>
<th>Total exon length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ex1</td>
<td>Ex2</td>
<td>Ex3</td>
</tr>
<tr>
<td>C.f. 37592</td>
<td>37592</td>
<td>15480</td>
<td>96</td>
</tr>
<tr>
<td>B.t. 34934</td>
<td>34934</td>
<td>18693</td>
<td>90</td>
</tr>
<tr>
<td>H.s. 32063</td>
<td>32063</td>
<td>11333</td>
<td>90</td>
</tr>
<tr>
<td>P.t. 32050</td>
<td>32050</td>
<td>11316</td>
<td>90</td>
</tr>
<tr>
<td>R.n. 22898</td>
<td>22898</td>
<td>9430</td>
<td>83</td>
</tr>
<tr>
<td>M.m. 22786</td>
<td>22786</td>
<td>9483</td>
<td>91</td>
</tr>
<tr>
<td>M.r. 21161</td>
<td>21161</td>
<td>7939</td>
<td>84</td>
</tr>
<tr>
<td>anc. 30297</td>
<td>30297</td>
<td>11506</td>
<td>93</td>
</tr>
<tr>
<td>anc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Exons (Ex) and pseudoexons (pEx). Species: Canis familiaris (C.f.), Bos taurus (B.t.), Homo sapiens (H.s), Pan troglodytes (P.t.), Rattus norvegicus (R.n.), Mus musculus (M.m.), Microtus rossiaemeriodianalis (M.r.) and ancestor (anc); (–) – the exon is missing.

The most variable among all the exons was exon 1, for its length can vary twofold (cow-vole). This fact could be explained by the presence of tandem blocks of repeats (repeats A-F) within exon 1 (not shown). Some repeats (A, F, B, D) within exon 1, conserved and variously amplified, occur in all species; others are specific for a certain species (the С repeat for mouse, the В repeat for man, the G repeat for cow). These repeats comprise 40 to 78% of the entire sequence, as in the case of bovine ex-1, in which we have identified a new tandem repeat, 10 kb in length, which we denoted as the G repeat.

What additionally contribute to the variability of the exon 1 length are insertions of mobile elements, that is, SINEs, which, again, are species specific, except for MIR family members, which are common to all the mammals (Fig. 1).

All the other exons display slight variation in length (Table 1). The most conserved is exon 4 (Ex5a - ancestral), which contains a long inverted repeat; however, the function of this exon within gene RNA is not yet clear.

As early as when it existed, the ancestral gene contained the basic domains of tandem repeats, A-E, which suggests that their amplification occurred more than 100 Myr ago; whereas the amplification of the species specific tandem repeats and insertion of the species specific SINEs was confined to species radiation or earlier times. It is most likely that the presence of species specific sequences (exons, repeats, mobile elements) in the Xist structure reflects the way in which the gene function within the X chromosome to which it originally belongs, and apparently is a common feature to all the regulatory genes of that type.

Based on the presence of the remains of ancient mobile elements (L1MC3, L2, L3, MIR) in the exons and introns of the ancestral gene, their high level of divergence (over 30%) and the results of comparison of the time of their active genome-wide spreading (about 200 Myr ago), we have concluded that the Xist gene might arise de novo during the first stage of mammalian sex chromosome formation: 240–320 Myr ago.
ACKNOWLEDGEMENTS

The work was supported in part by the Integration Projects of the Presidium of Russian Academy of Sciences (contract N10104-34/P-18/155-270/1105-06-001).

REFERENCES