

## Evolution of Chromatin-Remodeling Complexes: Comparative Genomics Reveals the Ancient Origin of “Novel” Compensasome Genes

Ignacio Marín

Departamento de Genética, Universidad de Valencia, Calle Doctor Moliner 50, Burjassot 46100, Spain

Received: 13 May 2002 / Accepted: 4 December 2002

**Abstract.** Dosage compensation in *Drosophila* is mediated by a complex, called compensasome, composed of at least five proteins and two noncoding RNAs. Genes encoding compensasome proteins have been collectively named *male-specific lethals* or *msls*. Recent work showed that three of the *Drosophila msls* (*msl-3*, *mof*, and *mle*) have an ancient origin. In this study, I describe likely orthologues of the two remaining *msls*, *msl-1* and *msl-2*, in several invertebrates and vertebrates. The MSL-2 protein is the only one found in *Drosophila* and vertebrate genomes that contains both a RING finger and a peculiar type of CXC domain, related to the one present in Enhancer of Zeste proteins. MSL-1 also contains two evolutionarily conserved domains: a leucine zipper and a second characteristic region, described here for the first time, which I have called the PEHE domain. These two domains are present in the likely orthologues of MSL-1 as well as in other genes in several invertebrate and vertebrate species. Although it cannot be excluded that the compensasome complex is a recent evolutionary novelty, these results show that all *msls* are found in mammals, suggesting that protein complexes related to the compensasome may be present in mammalian species. Metazoans that lack several of the *msls*, such as *Caenorhabditis elegans*, cannot contain compensasomes. The evolutionary relationships of the compensasome and the NuA4 complex, another chromatin-remodeling complex that contains related subunits, are discussed.

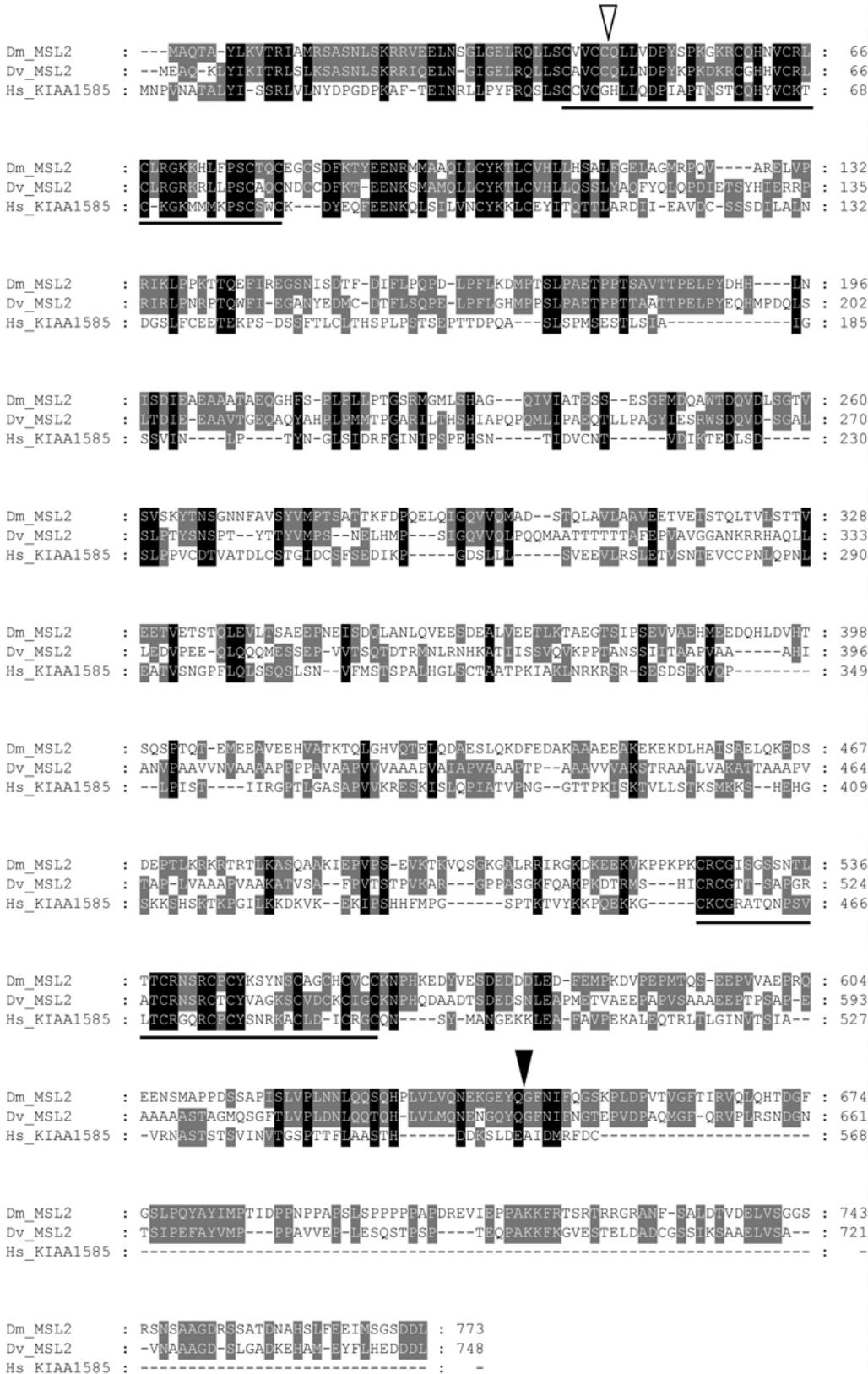
**Key words:** Dosage compensation — Chromatin — Enhancer of Zeste — Comparative genomics

### Introduction

In drosophilid species, dosage compensation is achieved by doubling the transcription of the genes of the single male X chromosome. This process involves a general modification of the chromatin structure of the X chromosome in males (reviewed by Baker et al. 1994; Cline and Meyer 1996). Dosage compensation is mediated, for most X chromosome genes, by the action of a complex that contains at least five proteins, encoded by the *male-specific lethal* genes (*male-specific lethal-1*, *-2*, and *-3* [*msl-1*, *msl-2*, *msl-3*], *maleless* [*mle*], and *males-absent-on-the-first* [*mof*]) and two noncoding RNAs (encoded by the *rox1* and *rox2* genes). This complex has been named the “compensasome” (reviewed by Franke and Baker 2000; Kelley and Kuroda 2000). A sixth protein, JIL-1, may also be part of the compensasome (Jin et al. 2000). How the compensasome determines hypertranscription on the X chromosome is still incompletely understood. It is, however, accepted that acetylation of histone H4 at its lysine 16 by MOF, which is a histone acetyltransferase of the MYST family (Hilfiker et al. 1997), is a significant feature of compensasome action (Turner et al. 1992; Bone et al. 1994; Smith et al. 2000).

The origin of the compensasome-based mechanism of dosage compensation is still obscure (Marín et al. 2000). Direct confirmation for the action of compensasomes in other organisms has been obtained for

A



**Fig. 1. A** Alignment of the sequences of the *Drosophila melanogaster* (Dm) and *Drosophila virilis* (Dv) MSL-2 proteins with *Homo sapiens* (Hs) KIAA1585. The underlined motifs correspond to the RING finger (close to the N terminus) and the CXC domain (toward the C-terminal end). The arrowheads point toward the positions of introns in *Drosophila* MSL-2 (filled arrowhead) and human KIAA1585 (Open arrowhead). In this and the following

figures, shading is used to highlight the presence of identical or biochemically related amino acids in a given position. **B** N-terminal ends of the *Drosophila* MSL-2 and human KIAA1585 proteins can be readily aligned with the KIAA1585-related sequence found in *Ciona intestinalis* (Ci) and with the *Anopheles gambiae* MSL-2 sequence (Ag; see text). The RING finger domain is underlined.

## B

```

Dm_MSL-2      : ---MAQTA-YLKVTRIAMRSASNLSKRRVEELNSGIGELRQLLS : 40
Dv_MSL-2      : --MEAQK-LYIKITRSLKKSASNLSKRRIQELN-GIGELRQLLS : 40
Anopheles_MSL-2 : MNP---TSLYITTSRLVFQ---GDIVAGLQDLNRLLPYLRKKSLS : 38
Hs_KIAA1585   : MNPVNATALYI-SSRLVFNYPGDPKAF-TEINRLLPYFRQSLS : 42
Ciona_AV6741258 : MDTENCNLYLQVCQTFELTCKYDDPSTF-EDLYKLLPWLRSLS : 43

Dm_MSL-2      : GVVCCQLLVDPYSPKGRKCOHNVCRCLGRGKHLF-PSCIQCEG : 83
Dv_MSL-2      : CAVCCQLLNDPYKPKDKRCGHHVCRCLGRGRKRL-LPSCAQCNL : 83
Anopheles_MSL-2 : CAVCCQLLVDPYSPKGRKCOHNVCRCLGRGKHLF-PSCIQCEG : 80
Hs_KIAA1585   : CCVCGHLLQDPIAPNSTCOHYVCKTC-KGKKMMKPSCSWCK- : 84
Ciona_AV6741258 : CCVCSKVVVNPMSSTQTHCOHFVCTKC-RGGKMLVLRPSCSWCK- : 85

Dm_MSL-2      : CSDFKTYEENRMMAAQLCYKTLCVHLLHSALFGEIAGMRPQV- : 126
Dv_MSL-2      : CCDFKT-EENKSMAMQLCYKTLCVHLLQSSLYAQFYQLQPDIE : 126
Anopheles_MSL-2 : --GLLQYRENTQLRLLCYKSICTFIRTRPVYADI--CKCAG- : 119
Hs_KIAA1585   : --DYEQFEENKQLSILVNCYKKLCEYITQTTLARDII-EAVDC- : 124
Ciona_AV6741258 : --DANAFEINEHLVILVOCFSKICELILNSPINDHI--IKYDQ- : 124

Dm_MSL-2      : ---ARELVPRIKLPPKTTQEFFIREGSNISDTF-DIFLPQPD-LP : 165
Dv_MSL-2      : TSYHIERRPRIRLNRPTQWFI-EGANYEDMC-DTFLSQPE-LP : 167
Anopheles_MSL-2 : ---GAGQPREGGGPGPSAASPHNTATTSPN-SLMELIEEGAA : 158
Hs_KIAA1585   : SSDILALLNDGSLFCEETKPS-DSSFTLCLTHSPHPSTSEPTT : 167
Ciona_AV6741258 : -NRKKEQQKEATGKKEEQSKRRSGTRRHQVS-SRLINISV-SE : 165

Dm_MSL-2      : FLKD : 169
Dv_MSL-2      : FLGH : 171
Anopheles_MSL-2 : FVDD : 162
Hs_KIAA1585   : DPOA : 171
Ciona_AV6741258 : VLQD : 169

```

Fig. 1. Continued

species of at least two genera of the *Drosophilidae* family (Marín et al. 1996), showing that this protein complex has been acting on dosage compensation for at least 50 million years. Beyond that phylogenetic range, the search for a similar mechanism in other species has been so far fruitless. It is well known that the nematode *Caenorhabditis elegans*, member of a lineage that diverged from the *Drosophila* lineage more than 500 million years ago, has a totally unrelated mechanism of dosage compensation (reviewed by Meyer 2000). Moreover, evidence that one of the proteins of the compensasome, MLE, is not acting on dosage compensation in another dipteran species, *Sciara ocellaris*, has been obtained (Ruiz et al. 2000).

Two lines of evidence suggest, however, that complexes that include at least some of the compensasome proteins are ancient. First, my coworkers and I showed that likely orthologues of three of the compensasome genes (*msh-3*, *msh-1*, and *mof*) exist in many organisms, particularly in mammals (Marín and Baker 2000; Sanjuán and Marín 2001). Second, it has been found that a *Saccharomyces cerevisiae* protein complex

known as NuA4 contains the protein encoded by the closest relative of *msh-3* in yeasts (that has been called *Alp13* or *Eaf3*), together with a relative of *mof*, called *Esal* (Allard et al. 1999; Eisen et al. 2001). These findings show that the interaction among *mof*-like and *msh-3*-like proteins is at least 1 billion years old.

Previous works on three compensasome genes (Marín and Baker 2000; Sanjuán and Marín 2001) revealed species in which some of those genes are missing (e.g., *C. elegans*) and also in which organisms, apart from *Drosophila*, all the genes are present (e.g., mammals). Based on these data, we proposed that protein complexes similar to the compensasome may exist in mammalian species. However, a serious problem with that hypothesis was the lack of evidence for the existence of orthologues of two of the compensasome genes, *msh-1* and *msh-2*, in mammals. In this work, based on recent data generated by genome sequencing projects, I describe likely orthologues of *msh-1* and *msh-2* in many species, including several vertebrates. Furthermore, I present evidence that MSL-2 contains a protein domain similar to the CXC domain of En-

**A**

```

Dm_MSL-2      : CRCGISGSSNTLTT CRNSRCP CYKSYNSC-AGCHCVCC
Dv_MSL-2      : CRCGT-TSAPGRAT CRNSRCT CYVAGKSC-VDCCKIGC
Hs/Mm_KIAA1585 : CKCGRATQNPVSLT CRGQRCP CYSNRKAC-LDCICRGC
Xl_BF025273   : CKCGRATQHPSVLT CRGQRCP CYSNRKAC-LDCICRGC
Anopheles_MSL-2 : CRCGNATPTPGKLT CCGQRCP CYVDSKSC-VDCCKRGC
Apis_MSL-2     : CRCGNATAIPGKLT CCGQRCP CYVESKPC-VECRCRGC
Ce_Mes2        : CNC-AAGQ-----CYTKACQ CYRANWEC-NPMTCNMC
Dm_E(z)        : CRC--KAQ-----CNTKQCP CYLAVREC-DPDLCCQAC
Hs_ENX-1       : CRC--KAQ-----CNTKQCP CYLAVREC-DPDLCLTC
Hs_ENX-2       : CRC--KTQ-----CNTKQCP CYLAVREC-DPDLCLTC
At_MEDEA       : CNC-AIGQ-----CTNRQCP CFAANREC-DPDLCRSC
At_EZA1        : CHC-AKSQ-----CRSRQCP CFAAGREC-DPDVCRNC
At_Curlyleaf   : CHC-AKSQ-----CRSRQCP CFAADREC-DPDVCRNC
Dm_CG14016     : CCC-KRSQ-----CIKNYCD CYQSMACITKFCRCVGC
At_F19B15.30a  : CNC-KHSR-----CLKLYCE CFASGYC-DGCNCVNC
At_TS01a       : CNC-KKSK-----CLKLYCE CFAAGFYCIEPCSCINC
At_MSL1.20a    : CRC-KQSK-----CLKLYCD CFASGVVC-TDCDCVDC
Dm_CG6061a     : CNC-SKSQ-----CLKLYCD CFANGFC-QDCTCKDC
Ce_JC8.6a      : CNC-TKSQ-----CLKLYCD CFANGFC-RDCNCKDC
At_AC009465    : CNC-RKSG-----CSKKYCE CFMMGVGCSSNCRCMGC
At_TS01-likeb  : CNC-KKSN-----CLKKYCE CYQGGVGC SINCRCEGC
At_TS01b       : CNC-KKSN-----CMKKYCE CYQGGVGC SMNCRCEGC
Ce_JC8.6b      : CHC-KKSG-----CLKNYCE CYEAKVPTDRCKCKGC
Dm_CG6061b     : CNC-KRSG-----CLKNYCE CYEAKIPCSSICKCVGC
Hs_Tesmin      : CNC-RRSG-----CLKNYCE CYEAQIMCSSICKCIGC
At_F19B15.30b : CHC-KKSG-----CLKKYCE CFQANILCSENCKCLDC
At_MSL1.20b    : CKC-KRTR-----CLKKYCE CFQANLLCSDNCKCINC

```

**Fig. 2. A** Alignment of CXC domains. Sequences are ordered according to the tree shown in **B**, clockwise, starting with Dm MSL-2. This alignment includes the whole sequences of the MSL-2 and TSO-1 CXC domains and a partial sequence of the E(Z) CXC domains, extended to include only the C-X-C triad that is closest to the [C-(X)<sub>4</sub>-C-X-C-(X)<sub>6</sub>-C-(X)<sub>4-5</sub>-C (X)<sub>2</sub>-C] fundamental sequence. E(Z) CXC domains can be extended farther N-terminally, to include two additional C-X-C motifs (see text). Abbreviations as follows: Hs, *Homo sapiens*; Mm, *Mus musculus*; Xl, *Xenopus*

hancer of Zeste (E[Z]) proteins. MSL-1 proteins contain a novel characteristic domain which I have called the PEHE domain. Several PEHE domain-containing genes are found in both vertebrates and invertebrates.

## Materials and Methods

**Data Mining.** Database searches were performed online and finished in early May 2002. All data presented were obtained in searches performed using the PSI-BLAST, BLASTP, or TBLASTN program (Altschul et al. 1997) and the “nonredundant,” “month,” “dBEST,” “htgs,” “wgs\_Anopheles,” “human genome,” or “unfinished microbial” database at the National Center for Biotechnology Information (NCBI) web pages (<http://www.ncbi.nlm.nih.gov/>).

**Sequence Alignments.** Multiple-sequence alignments were obtained using the default parameters of the ClustalX program (version 1.81 [Thompson et al. 1997]) and refined manually, using GeneDoc 2.5 (Nicholas and Nicholas 1997) to edit the sequences.

*laevis*; Dm, *Drosophila melanogaster*; Dv, *Drosophila virilis*; At, *Arabidopsis thaliana*; Ce, *Caenorhabditis elegans*. The CXC domains of human and mouse KIAA1585 are identical. A lowercase letter (a, b) has been added to distinguish the two TSO-1 CXC domains present in several of these proteins. The first TSO-1 CXC domain of *Arabidopsis thaliana TSO1-like* is identical to that found in *TSO-1* and it is not included. **B** Phylogenetic tree showing the relationships among the sequences presented above. Only bootstrap values over 600 are detailed.

GeneDoc was also used to draw and highlight the similarities among sequences in the figures presented here. Pairs of sequences were aligned online using BLAST 2 (Tatusova and Madden 1999), at the corresponding NCBI web page (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/bl2.html>).

**Phylogenetic Trees.** The neighbor-joining (Saitou and Nei 1987) routine implemented in ClustalX was used to establish phylogenetic relationships among sequences (see below). A total of 1000 bootstrap replicates was performed to determine the reliability of the tree topology. TreeView 1.6.1 (Page 1996) was used to draw the trees shown in the next section.

## Results

### *An Orthologue of the Male-Specific Lethal-2 Gene is Present in Vertebrates*

Database searches established that a human protein, called KIAA1585 (Nagase et al. 2000), was substan-

B

**TSO-1 CXC domains**

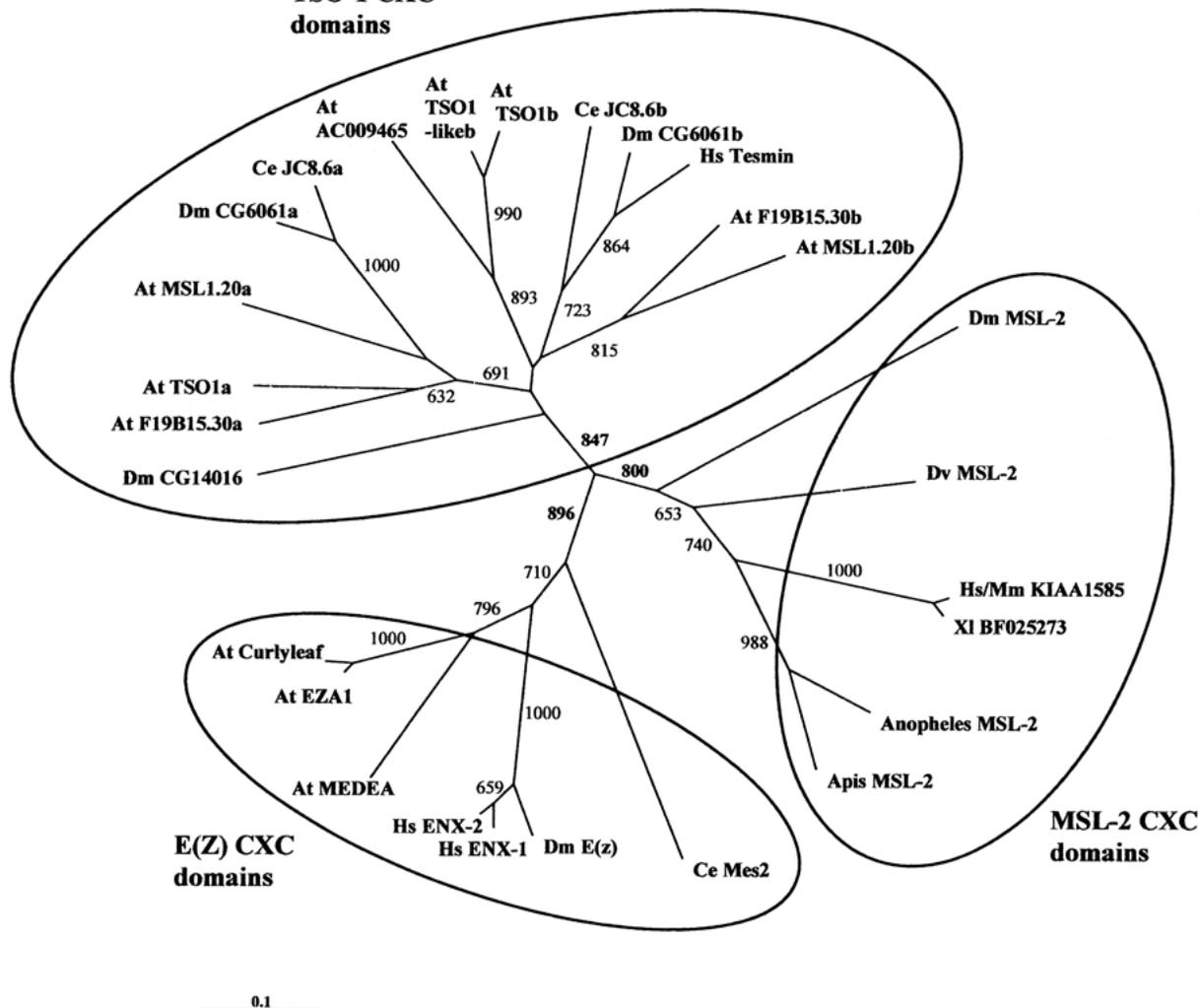


Fig. 2. Continued

tially similar to *D. melanogaster* MSL-2. The “expect value” (*E* value), i.e., the number of proteins in the database expected to have that level of similarity by chance, was  $3 \times 10^{-6}$ . The accession numbers for KIAA1585 are, respectively, AB046805 for the cDNA and AC026969 for the corresponding genomic DNA sequence, derived from chromosome 3. Figure 1A shows the alignment of this incompletely characterized human protein with the *D. melanogaster* (Bashaw and Baker 1995; Zhou et al. 1995) and *D. virilis* (Coppes et al. 1998) MSL-2 proteins. Analyses of the available cDNAs did not allow us to exclude the possibility that the N terminus of the human protein extends farther upstream of the methionine amino acid shown as the starting codon in Fig. 1A. As shown in Fig. 1A, the most similar regions in MSL-2 and KIAA1585 include the two cysteine-rich domains described in MSL-2 (Zhou et al. 1995; Bashaw and Baker 1995). They are an N-terminal RING finger and a second Cys-rich domain, located C-terminally, which was previously described

as a truncated PHD domain (Bashaw and Baker 1995) or as a divergent metallothionein-like domain (Zhou et al. 1995). Exhaustive database searches determined that no other available sequence in any species can encode a protein containing those two domains at the same time, a result that strongly suggests that the genes are orthologous. Moreover, PSI-BLAST analyses using both the RING finger and the second Cys-rich domain in MSL-2 determined that the closest relative to the *Drosophila* protein was in both cases KIAA1585, giving further support to the hypothesis that the genes are orthologous. Therefore, I call *Hs msl-2* the gene encoding KIAA1585.

Partial EST sequences encoding fragments of proteins almost identical to the human one exist in other vertebrate species, such as the mammals *Mus musculus* (accession number BF471233), *Rattus norvegicus* (AF202267), and *Bos taurus* (AW462003), the bird *Gallus gallus* (AI981288), the amphibian *Xenopus laevis* (BF025273), and the fish *Danio rerio*

(AI657716). Vertebrate *msl-2* cDNAs were detected in libraries obtained from many different tissues or developmental stages. Thus, *Hs msl-2* cDNAs were obtained from colon (AW752909), stomach (BF374-080), lung (AW139216), T lymphocytes (AA352974), or placenta (AU136838). Mouse cDNAs were derived from two-cell eggs (BB290571), embryos (AA387-285), thymus (AA863997), or spleen (BE849962). A single cow cDNA of known origin was derived from placenta (AW462003). Another cDNA, this time from chicken, came from T lymphocytes (AI981288). Finally, a related cDNA obtained from the toad *Xenopus laevis* was derived from gastrulae (BF025273).

Interestingly, cDNAs of the urochordate *Ciona intestinalis* (accession numbers AV674158 and AV671862) also showed a very high similarity to the human *msl-2* gene ( $E$  value =  $10^{-27}$ ). However, the two *Ciona* cDNAs encode a very short protein, of only 181 amino acids. The putative protein would contain the RING finger but not the second Cys-rich domain. Whether this is indeed a full-length protein or simply a truncated product cannot be determined with the available data. Additionally, I detected a region with a high similarity to *D. melanogaster msl-2* in a sequence very recently obtained in the genome sequencing project of the mosquito *Anopheles gambiae* (accession number AA-AB01008841.1). A putative protein of 1024 amino acids is encoded in nucleotides 1760353–1763427 of this sequence that also contains the two characteristic cysteine-rich domains present in MSL-2 and KIAA1585 and, thus, must correspond to the mosquito *msl-2* gene. Figure 1B shows an alignment of the N terminus of the *Ciona* and *Anopheles* sequences with those of the *Drosophila* MSL-2 and human KIAA1585 proteins.

#### *The Second Cysteine-Rich Motif in MSL-2 is a Divergent CXC Domain*

Comparisons of the *D. melanogaster* and *D. virilis* MSL-2 proteins with mammalian KIAA1585 are useful to determine the pattern of conserved residues characteristic of the C-terminal cysteine-rich domain. It can be generalized as  $[C-X-C-(X)_{10-11}-C-(X)_4-C-X-C-(X)_6-C-(X)_2-C-X-C-(X)_2-C]$ , a formula that I refer to, for reasons that will become immediately clear, as the “MSL-2 CXC domain.” When the sequences of the MSL-2 CXC domain of any of these three species were used to perform highly sensitive TBLASTN and PSI-BLAST searches, I found only two other sequences with such a formula, derived from insects (*Anopheles gambiae* and *Apis mellifera*, respectively). However, other proteins in the databases showed related domains with very similar

spacings of cysteine residues. They corresponded to two groups of proteins. On one hand, I detected the Enhancer of Zeste [E(Z)] proteins of several animal and plant species (Jones and Gelbart 1993; Abel et al. 1996; Chen et al. 1996; Hobert et al. 1996a, b; Goodrich et al. 1997; Grossniklaus et al. 1998; Holdeman et al. 1998; Cardoso et al. 2000). They have a long cysteine-rich region known as the “CXC domain” (Hobert et al. 1996a), which I call more precisely the “E(Z) CXC domain.” On the other hand, several proteins have one or, more often, two cysteine-rich domains that are slightly shorter than the one found in MSL-2 proteins. Two groups that detected these regions, repeated twice in the *Arabidopsis thaliana* TSO1 gene, called them “TCR repeats” (Song et al. 2000) or, alternatively, CXC domains (Hauser et al. 2000), because of their similarity to the E(Z) CXC domains. I call these domains “TSO-1 CXC domains.” The reason to call these three cysteine-rich regions CXC domains is simple. For these three groups of proteins, the C-terminal end of the domains is indeed very similar and, also, exclusive for this group of proteins. It has the formula  $[C-(X)_4-C-(X)-C-(X)_6-C-X_{4-5}-C-(X)_2-C]$ . Moreover, although the N terminus of these domains is different in the three groups, it always includes one to three copies of a C-X-C motif. Thus, they can be considered three related types of CXC domains with different N-terminal extensions.

The alignment of the similar C terminus of representative proteins containing these three CXC domains is shown in Fig. 2A. A phylogenetic tree built according to that alignment is shown in Fig. 2B. Assuming that these domains have a common origin (instead of having related sequences as a by-product of evolutionary convergence), their relationships can be established as follows. The topology of the tree is compatible with the three types of domains emerging and becoming differentiated before the animal/plant split, with a subsequent loss of the MSL-2 class of CXC domains in plants. However, an equally parsimonious alternative is an ancient origin for the E(Z) and TSO1 CXC domains, with the MSL-2 CXC domain being a later derivative of one of these two, arisen in animals. The tree also suggests that proteins with double TSO1 CXC domains emerged from a protein with a single TSO1 CXC domain, in which the domain became duplicated. That duplication also occurred before the animal/plant split, but when the TSO1 CXC domains were already very different from the rest of the CXC domains. Very interestingly, no known yeast or fungal protein has CXC domains, although the existence of both the TSO1 and the E(Z) CXC domains in plants and animals suggests that at least these two domains must have existed in the common ancestor of all fungi.



## A

```

Dm_MSL-1 : SLEDFKARWVP----- : 11
Mm_MSL-1 : VLAWPSWRDHS----- : 11
Hs_MSL-1 : TVAWPSWRDHS----- : 11
Xl_MSL-1 : SVAWPSWEHR----- : 11
Ag_MSL-1 : SLEDFSWTVIEDDG---- : 14
As_BM283183 : RLEDFPFRNVGR----- : 12
Bm_AA471573 : RVECFSEFRKVHD----- : 12
Na_BG467879 : QIQDFEVKVVGR----- : 12
Mm_KIAA1267 : EILTFPSARWVD----- : 11
Hs_KIAA1267 : EILTFPSARWVD----- : 11
Xl_BJ072550 : EILTFPSRWVLDLVE---- : 14
Mm_BG298804 : EILTFPSRRKVV----- : 11
Hs_AL133053 : EILTFPSRWVV----- : 11
Dm_CG4699 : EIPDFKRRIVDSNDKDGKHSDESMERKLSNGCVASKSEELAKQQLPPPPDQDQEQPALQPPKVNGLKEN : 70
Ag_BM619455 : DIAVFKRRRIVDQYQENGDE----SLTQMDSSATANELTIGMELSTEENPTTTPVKVVDGA----- : 55

```

```

Dm_MSL-1 : ----- : -
Mm_MSL-1 : ----- : -
Hs_MSL-1 : ----- : -
Xl_MSL-1 : ----- : -
Ag_MSL-1 : ----- : -
As_BM283183 : ----- : -
Bm_AA471573 : ----- : -
Na_BG467879 : ----- : -
Mm_KIAA1267 : ----- : -
Hs_KIAA1267 : ----- : -
Xl_BJ072550 : ----- : -
Mm_BG298804 : ----- : -
Hs_AL133053 : ----- : -
Dm_CG4699 : QAVKHNNNNNTNNNNKGLVNGNAKDELKAAEDHDTQSKEDDDPKKVVVKPLNGNLIKNPIDKTAET : 140
Ag_BM619455 : ----- : -

```

```

Dm_MSL-1 : -----VPE--SSDKID-----TEELSPATERRRHKYVKD : 39
Mm_MSL-1 : -----VEPLRDPNPSDI-----TENLDPSVSRHRHAKLELD : 42
Hs_MSL-1 : -----VEPLRDPNPSDL-----TENLDPSVSRHRHAKLELD : 42
Xl_MSL-1 : -----TEPLQVTELSDI-----PENLEDAVRAKRHAKLELD : 42
Ag_MSL-1 : -----DDAGGCDPPGGGEEEPSVSASE-----PSRENISDEAYAKRHTKLEID : 58
As_BM283183 : -----MEPLRSTPRLN-----DEVCTDEAYAKRHKYRQ : 42
Bm_AA471573 : -----LEPLRATPRHN-----DEVCTDEAYAKRHKYRQ : 42
Na_BG467879 : -----LSALRSQPRLT-----DERVDVDTVVRHREKLEKE : 42
Mm_KIAA1267 : -----VLSLKGSPDEENE-----EMEDLSAARAAALHAKCEM : 44
Hs_KIAA1267 : -----LSLKGSPDEENE-----EMEDLSAARAAALHAKCEM : 44
Xl_BJ072550 : -----RFLYGMDFPQ-----LEDLDEAYAKRHSAYBEL : 43
Mm_BG298804 : -----LOPLDEHN-LNKE-----EMEDLSDEVSRHRHAKYBER : 43
Hs_AL133053 : -----LOPLDEYN-LGKE-----EMEDLSDEVSRHRHAKYBER : 43
Dm_CG4699 : QDKIAQPAPEPPLPKRPKLETPTSNEVTKANGQLVQQLESRESEEPQHVDEDLSEAVTMRHRAALLE : 210
Ag_BM619455 : -----LPENGTKTTPSATIVPTTGVEAAYQQQLLIVIPAQ-----DEEDLSEETIMLKHRAALQE : 111

```

```

Dm_MSL-1 : FVDRKCRDARYMKEQIRLEQIRMRNQD-EVLVALDPLRASTFYF--LPE-DIEATQFVNEVTQVAF-- : 102
Mm_MSL-1 : EKRRKRWDIQRIRBORLQRLQLRMYKK--KGNQSESEPEVVSFFP--EPD-DVESLITTEFLPVVAE-- : 104
Hs_MSL-1 : EKRRKRWDIQRIRBORLQRLQLRMYKK--KGNQSESEPEVVSFFP--EPD-DVESLITTEFLPVVAE-- : 104
Xl_MSL-1 : EKRRKRWDIQRIRBORLQRLQLRMCKK--KGNQSESEPEVVSFFP--DVD-DVESVMITESLPPVVV-- : 104
Ag_MSL-1 : ERRKRWVQVIRBORHIREKRRKRLKE-QPVEQBAKAIITLYF--TVD-TLKYMLVTDVDPVQAF-- : 121
As_BM283183 : EKSIRRRDRIWQREEQYRLHL-MKISTDE----- : 70
Bm_AA471573 : EKLVKRRDRIQREEQYRLHL-MKVEKSD----- : 70
Na_BG467879 : EKQLRRDRIWQNEVYRSL-LGLQ----- : 67
Mm_KIAA1267 : ER-ARLWTTSPVPPRRGSSRSYSSDGRTPQLGSANP--STPQF--ASPD--VSSSHSLSEFSGHQSPR : 107
Hs_KIAA1267 : ER-ARLWTTSPVPPRRGSSRSYSSDGRTPQLGSANP--STPQF--ASPD--VSSSHSLSEYSGHQSPR : 107
Xl_BJ072550 : ER-SRWDFWSTAATRRGSSSNKSDGRWTPQPSNSQ--GSLQF--ASPDIGFHSNDVFTLSSAAPD : 108
Mm_BG298804 : EQ-ARNSLWEQSKWHRNRNRAYSKRVGQDLLKHSSELGSA---QQG-TAESPFLEAESHSILC-- : 104
Hs_AL133053 : EQ-ARNSLWEQSKWHRNRNRAYSKRVGQDLLKHPNPFSS---QQC-AAASPPGLPSENQDLC-- : 104
Dm_CG4699 : ER-RRFETFLKFPWSTR-SRANRRVDSR-AESSGANTDPASPAPHLGGPGHDNESTPSPLAYPLDAENE : 277
Ag_BM619455 : ER-KKERSVLKFPWSE--PRAKRRVDSR--AGSSGGNTDPTSPAP--PTPMVDHDSFPACESTPLTPDQG : 175

```

```

Dm_MSL-1 : -----GENVNM--EARDDFG-VPW : 119
Mm_MSL-1 : -----GRPLPKL--APFNFE-LPW : 120
Hs_MSL-1 : -----GRPLPKL--TPFNFE-LPW : 120
Xl_MSL-1 : -----GRPLPKL--TPFNFE-LPW : 120
Ag_MSL-1 : -----GELPLPL--PUNGGSF-LPW : 138
As_BM283183 : -----DPEPKV--QVVEVVSRAEK : 88
Bm_AA471573 : -----ASPEPKI--ESIEVVEKKEK : 88
Na_BG467879 : -----QTPEPLR--DFRIVVMKAR : 85
Mm_KIAA1267 : SPISPELHSAPL---TPVARDLRLHLASED---TRCSTPEL-GLDEGSV--CPW : 152
Hs_KIAA1267 : SPISPELHSAPL---TPVARDLRLHLASED---TRCSTPEL-GLDEGSV--CPW : 152
Xl_BJ072550 : SPEPYSTVQC-----LSVRSSTTRGLWSWSED---TRSSTPEI-PDEEIQIV-LPW : 152
Mm_BG298804 : -----AQDSLNL-NDGQEDKS-LRW : 122
Hs_AL133053 : -----AYGLFSL-NQSCETKS-LRW : 122
Dm_CG4699 : SGELLTGGQA-----RQARRRSTSSKLDQLERRSTTEPL-RESYTVMP-SPE : 323
Ag_BM619455 : QELSEANVMNALNKLHKKERRRVTSSKREDQ--HAQGAGITPTDTRVVV--SPY : 226

```

**Fig. 4. A** Alignment of the 15 sequences containing PEHE domains. Both the *D. melanogaster* gene *CG4699* and a very similar *Anopheles* sequence (Ag\_BM619455) contain a long string of amino acids in the middle of this domain. Sequences ordered according to **B**, counterclockwise starting with DmMSL1. **B** Phylogenetic tree obtained for the sequences shown in A.

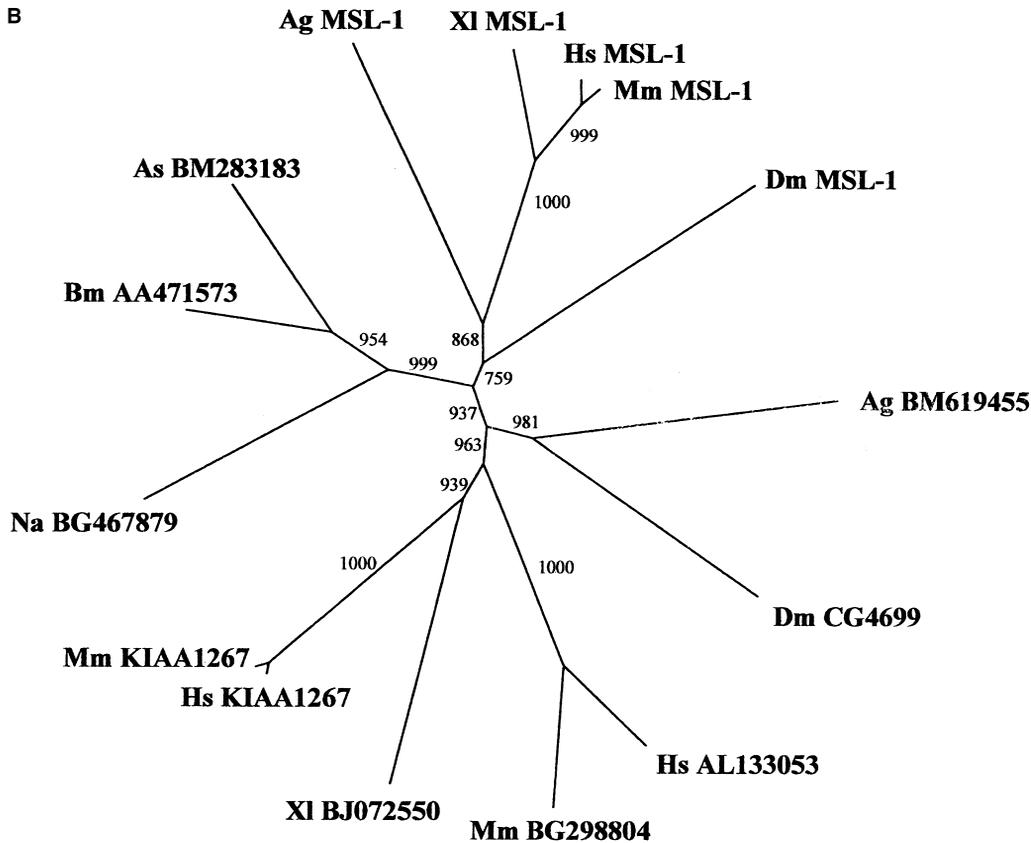


Fig. 4. Continued

*mssl-1 Shows a High Similarity to Other Invertebrate and Vertebrate Genes*

The only characteristic region hitherto detected in MSL-1 protein was a quite nonspecific acidic domain (Palmer et al. 1993). However, I found that MSL-1 contains two other domains that, for the first time, specifically relate *mssl-1* to other invertebrate and vertebrate genes. TBLASTN searches against the nonredundant database using MSL-1 as query detected a very significant similarity (*E* value,  $6 \times 10^{-9}$ ; 55% similarity over 123 residues) to a mouse cDNA (accession number AK014463) that encodes an apparently full-length 616-amino acid protein. The similarity extended from amino acid 866 to amino acid 985 in MSL-1, that is, outside the previously described acidic region. BLAST 2 analyses were used to confirm this finding. They revealed two regions of similarity between *Drosophila* MSL-1 and the protein putatively encoded by the mouse cDNA. First, BLAST 2 detects a 253-amino acid region, which includes the just mentioned conserved domain, extending from amino acid 748 to amino acid 985 in MSL-1 (corresponding to amino acids 351–595 in the

putative mouse protein; 25% identity, 44% similarity). A second region was detected by BLAST2 analyses extending from amino acid 81 to amino acid 165 in MSL-1, which has 21% identity and 52% similarity to amino acids 181–279 of the putative mouse protein. This second region corresponds to the putative leucine zipper domain described by Marin and Baker (2000). Systematic PSI-BLAST and TBLASTN searches demonstrated that several other genes have similar protein domains. Figure 3 shows the alignment of the *Drosophila* MSL-1 protein with those encoded by the genes most similar to MSL-1 found in the databases, which derive from the genomes of the mosquito *Anopheles gambiae*, the mammals *Mus musculus* and *Homo sapiens* and the amphibian *Xenopus laevis* (a partial sequence). The human sequence was reconstructed by comparison with the almost-identical mouse sequence, by aligning overlapping cDNAs (accession numbers AI800794, BG421221, AW965141, AV656786, and H86733) and the human genomic sequence NT\_010844.3, from chromosome 17. For reasons that will become immediately clear, I consider these genes to be orthologues of *mssl-1*, giving them the same name but using

two letters to refer to the particular species in which they are found. Interestingly, no acidic domain is detected in the vertebrate MSL-1 proteins, which are, moreover, quite shorter than their invertebrate counterparts. PAIRCOIL analyses (Berger et al. 1995) showed that the conserved N-terminal region found in all these proteins indeed corresponds to a coiled-coil domain, potentially a leucine zipper, strengthening the suggestion by Marín and Baker (2000) that this region must be significant for protein function.

The conserved C-terminal domain is found in other genes apart from those shown in Fig. 3. A total of 15 sequences was detected that had significant similarity in this region. They are two *Drosophila* genes (*mssl-1* and *CG4699*), two *Anopheles gambiae* genes (*Ag mssl-1* and a second, unknown gene), three mammalian genes (namely, *mssl-1*, a gene known as *KIAA1267* in humans, and a third, previously undescribed gene; both the mouse and the human genes are shown), two genes from *Xenopus laevis* (*Xl mssl-1* and a second gene, similar to *KIAA1267*), and a single gene from each of three nematode species, *Ascaris suum*, *Brugia malayi*, and *Necator americanus*. Interestingly, the fully sequenced nematode *Caenorhabditis elegans* does not seem to contain any gene with this domain. The alignment of all these sequences is shown in Fig. 4A. Because I found four characteristic amino acid residues (P, E, H, E) that are identical in all these sequences, I have called this conserved region the PEHE domain.

Figure 4B shows a phylogenetic tree obtained from the alignment in Fig. 4A. It demonstrates that *Drosophila mssl-1* is more similar to the genes that I also named *mssl-1* than to the other *Drosophila* gene (*CG4699*) or to any of the other sequences. I took advantage of the fact that apparently full-length sequences are available for *Drosophila* *CG4699* (1570 amino acids) and human *KIAA1267* (1105 amino acids) proteins to confirm the close relationship of these sequences, suggested by the alignment of their PEHE domains. BLAST2 analyses detected a 640-position-long region of similarity (corresponding to amino acids 560 to 1114 in the *Drosophila* *CG4699* protein and amino acids 355 to 929 in the human *KIAA1267* protein; 24% identity, 38% similarity).

The simplest interpretation of the results shown in Figs. 4A and B is that an ancestral gene was duplicated to give rise to two *mssl-1*-like genes prior to the protostome/deuterostome dichotomy. These genes differentiated substantially, but still conserved the characteristic PEHE domain. One of those genes became *mssl-1* in *Drosophila* and mammals. The second became *CG4699* in *Drosophila* but was again duplicated after the protostome/deuterostome split, to give rise to the two additional genes found in mammals. This is the evolutionary hypothesis that

suggests that the *Drosophila* and vertebrate *mssl-1* genes should be considered orthologous. Alternative hypotheses would require much less parsimonious scenarios (e.g., duplications of an *mssl-1* ancestor before the protostome/deuterostome split, followed by losses of a different duplicate in each descendant lineage).

cDNAs containing sequences almost identical to regions of *Mm* or *Hs mssl-1* are found in other mammals (*Rattus norvegicus* [accession number BF284375], *Bos taurus* [BF045095]), in birds (*Gallus gallus* [AJ395334]), and in fish (*Danio rerio* [BG308211]). *Hs mssl-1* is, similarly to *Hs mssl-2*, expressed in multiple tissues. The databases contain cDNAs obtained from mammary gland (accession number BE693636), ovary (AI954139), colon (AW854130), bone marrow (AV762388), pituitary gland (AV745136), testis (BF888145), brain (R17800), retina (H86733), and fetal lung (AA374037).

## Discussion

Comparative genomic analysis allows the precise determination of the origin of evolutionary novelties at the molecular level. In this and previous work (Marín and Baker 2000; Sanjuán and Marín 2001), we have established the most likely evolutionary origin of the five *male-specific lethal* genes. Here, I have shown that likely orthologues of the hitherto “novel” *mssl-1* and *mssl-2* genes exist in vertebrate species. Although the similarity between MSL-2 and KIAA1585 is limited, the finding that these are the only two proteins found so far that have both a MSL-2 CXC domain and a RING finger domain is good evidence of their being orthologues. It is thus not surprising that KIAA1585 contains the RING finger that is the most similar to the one in MSL-2. The same logic applies to MSL-1 and its relatives. Although, again, their similarity is restricted, the finding of two characteristic, conserved domains that are typical of this group of proteins suggests that they have a common origin. This similarity among orthologues limited to a few conserved modules is reminiscent of what happens in the MSL-3 family, where only five short conserved domains are found (Marín and Baker 2000).

The characterization of the MSL-2 CXC domain is significant. The contradictory explanations proposed for this domain so far (a truncated PHD domain or an uncommon methallothionein-like domain; see, respectively, Bashaw and Baker [1995] or Zhou et al. [1995]) were unsatisfactory. The finding of a substantial similarity of a region of MSL-2 with the CXC domains typical of other proteins involved in chromatin-associated complexes, such as Enhancer of Zeste proteins, is suggestive of related roles. Evidence

that the CXC domains are significant for the function of these proteins is available. Mutations causing single-amino acid changes in the CXC domains of either *Drosophila E(z)* or *Arabidopsis TSO1* have been shown to cause strong phenotypic effects (Carrington and Jones 1996; Song et al. 2000; Hauser et al. 2000). Moreover, directed mutagenesis of the CXC domain in MSL-2 led to delayed development and reduced viability of the mutant males (Lyman et al. 1997). However, the precise role of the CXC domains is unknown. Evidence is available suggesting that the compensasome is built in the cell starting with a precursor MSL-1/MSL-2 heterodimer (Lyman et al. 1997; Gu et al. 1998). Chang and Kuroda (1998) presented evidence for MSL-1 having the ability to associate nonspecifically to chromosomes, the X-chromosome specificity being provided by the interaction of MSL-1 with MSL-2. Thus, CXC domains may be favoring interactions with other proteins or contributing to the recognition of specific DNA sequences. However, CXC function cannot be understood in isolation from the rest of the protein. Particularly, the unique combination of a RING finger and a CXC domain that is characteristic of MSL-2 proteins may have been critical for the emergence of the compensasome. The RING finger is crucial for the MSL-1/MSL-2 interaction mentioned above (Coppes et al. 1998).

It has been argued recently that a protein complex found in yeast and mammals, called NuA4, and the *Drosophila* compensasome may be the same complex in different species. This hypothesis would be supported by the presence of "homologous" (i.e., orthologous) proteins in both complexes (Pannuti and Lucchesi 2000; Eisen et al. 2001). In particular, the products of the yeast genes *Eaf3* and *Esal*, related to the *Drosophila* genes *msh-3* and *mof*, respectively, have been characterized in the yeast NuA4 complex (Allard et al. 1999; Eisen et al. 2001). However, Sanjuán and Marín (2001) argued against such a view, which does not take into account gene duplications. There are two facts to consider. First, the closest *Drosophila* relatives to the yeast NuA4 complex proteins EAF3 and ESA1 are not in fact the compensasome proteins MSL-3 and MOF, but those encoded by two *Drosophila* paralogues of the compensasome genes (called *MRG-15* and *EG0007.7*, respectively). Second, the suggestion that MOF has never been part of NuA4 is supported by results derived from the partial characterization of a mammalian NuA4 complex (Ikura et al. 2000). If NuA4 and the compensasome were in origin the same complex, mammalian NuA4 should include the mammalian orthologue of *D. melanogaster mof*, known as *MOF1* (Neal et al. 2000; Sanjuán and Marín 2001). However, mammalian NuA4 complexes do not contain the MOF1 protein. They instead

contain the protein encoded by a different MYST family gene, *TIP60*, which is the mammalian orthologue of *S. cerevisiae Esal* and *Drosophila EG0007.7* and a paralogue of *Drosophila mof* and human *MOF1* (Sanjuán and Marín 2001). Actually, a very recent study has shown that the human MOF1 protein is found in a different complex, which includes MRG15 (Pardo et al. 2002). These results strongly suggest that, although the interaction between MSL3- and MOF-related proteins is ancient, duplications of these genes led to the generation of different complexes that contain both types of proteins. In particular, the mammalian NuA4 complex and the *Drosophila* compensasome contain products encoded by paralogous genes, and therefore, these two complexes may have arisen totally independently.

This study reinforces the hypothesis of an ancient origin for all compensasome genes. Our data (Marín and Baker 2000; Sanjuán and Marín 2001; this study) are compatible with *mof* emerging by gene duplication before the fungus/animal split. Two other genes (*msh-3* and *mle*) may also have emerged by gene duplication after that divergence. Both *msh-1* and *msh-2* may have originated by the generation of novel domains (e.g., the PEHE domain in *msh-1*) and the establishment of novel domain combinations (leucine zipper + PEHE in *msh-1* genes, RING + CXC in *msh-2* genes) in animals. In summary, four of the five *mshs* may be exclusive of metazoans. The evidence showing that all five *msh* genes exist in mammalian genomes suggests that compensasome-like complexes may have existed prior to the protostome/deuterostome dichotomy. While compensasome-related complexes may be present in mammals, some protostome species, such as the nematode *Caenorhabditis elegans*, cannot contain compensasomes, due to the fact that orthologues of *msh-3* (Marín and Baker 2000), *msh-1*, and *msh-2* (this work) are not found. However, the possibility of all compensasome proteins coalescing only very recently to give rise to this complex in a few drosophilid species cannot be excluded at present.

## References

- Abel KJ, Brody LC, Valdes JM, Erdos MR, McKinley DR, Castilla LH, Merajver SD, Couch FJ, Friedman LS, Ostermeyer EA, Lynch ED, King MC, Welch PL, Osborne-Lawrence S, Spillman M, Bowcock AM, Collins FS, Weber BL (1996) Characterization of *EZH1*, a human homolog of *Drosophila Enhancer of zeste* near *BRCAl*. *Genomics* 37:161–171
- Allard S, Utley RT, Savard J, Clarke A, Grant P, Brandl CJ, Pillus L, Workman JL, Côté J (1999) NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *EMBO J* 18:5108–5119
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402

- Baker BS, Gorman M, Marín I (1994) Dosage compensation in *Drosophila*. *Annu Rev Genet* 28:491–521
- Bashaw GJ, Baker BS (1995) The *msl-2* dosage compensation gene of *Drosophila* encodes a putative DNA-binding protein whose expression is sex specifically regulated by *Sex-lethal*. *Development* 121:33245–33258
- Berger B, Wilson DB, Wolf E, Tonchev T, Milla M, Kim PS (1995) Predicting coiled coils by use of pairwise residue correlations. *Proc Natl Acad Sci USA* 92:8259–8263
- Bone JR, Lavender J, Richman R, Palmer MJ, Turner BM, Kuroda MI (1994) Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila*. *Genes Dev* 8:96–104
- Cardoso C, Mignon C, Hetet G, Grandchamps B, Fontes M, Colleaux L (2000) The human *EZH2* gene: Genomic organisation and revised mapping in 7q35 within the critical region for malignant myeloid disorders. *Eur J Hum Genet* 8:174–180
- Carrington EA, Jones RS (1996) The *Drosophila Enhancer of zeste* gene encodes a chromosomal protein: Examination of wild-type and mutant protein distribution. *Development* 122:4073–4083
- Chang KA, Kuroda MI (1998) Modulation of MSL1 abundance in female *Drosophila* contributes to the sex specificity of dosage compensation. *Genetics* 150:699–709
- Chen H, Rossier C, Antonarakis SE (1996) Cloning of a human homolog of the *Drosophila enhancer of zeste* gene (*EZH2*) that maps to chromosome 21q22.2. *Genomics* 38:30–37
- Cline TW, Meyer BJ (1996) Vive la différence: Males vs. females in flies and worms. *Annu Rev Genet* 30:637–702
- Copps K, Richman R, Lyman LM, Chang KA, Rampersad-Ammons J, Kuroda MI (1998) Complex formation by the *Drosophila* MSL proteins: Role of the MSL2 RING finger in protein complex assembly. *EMBO J* 17:5409–5417
- Eisen A, Utley RT, Nourani A, Allard S, Schmidt P, Lane WS, Lucchesi JC, Côté J (2001) The yeast NuA4 and *Drosophila* MSL complexes contain homologous subunits important for transcriptional regulation. *J Biol Chem* 276:3484–3491
- Franke A, Baker BS (2000) Dosage compensation rox! *Curr Opin Cell Biol* 12:351–354
- Goodrich J, Puangsomlee P, Martin M, Long D, Meyerowitz EM, Coupland G (1997) A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386:44–51
- Grossniklaus U, Vielle-Calzada JP, Hoepfner MA, Gagliano WB (1998) Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in *Arabidopsis*. *Science* 280:446–450
- Gu W, Szauter P, Lucchesi JC (1998) Targeting of MOF, a putative histone acetyl transferase, to the X chromosome of *Drosophila melanogaster*. *Dev. Genet* 22:56–64
- Hauser BA, He JQ, Park SO, Gasser CS (2000) TSO1 is a novel protein that modulates cytokinesis and cell expansion in *Arabidopsis*. *Development* 127:2219–2228
- Hilfiker A, Hilfiker-Kleiner D, Pannuti A, Lucchesi JC (1997) *mof*, a putative acetyl transferase gene related to the *Tip60* and *MOZ* human genes and to the *SAS* genes of yeast, is required for dosage compensation in *Drosophila*. *EMBO J* 16:2054–2060
- Hobert O, Jallal B, Ullrich A (1996a) Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression. *Mol Cell Biol* 16:3066–3073
- Hobert O, Sures I, Ciossek T, Fuchs M, Ullrich A (1996b) Isolation and developmental expression analysis of *Enx-1*, a novel mouse Polycomb group gene. *Mech Dev* 55:171–184
- Holdeman R, Nehrt S, Strome S (1998) MES-2, a maternal protein essential for viability of the germline in *Caenorhabditis elegans*, is homologous to a *Drosophila* Polycomb group protein. *Development* 125:2457–2467
- Ikura T, Ogryzko VV, Grigoriev M, Groisman R, Wang J, Horikoshi M, Scully R, Qin J, Nakatani Y (2000) Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. *Cell* 102:463–473
- Jin Y, Wang Y, Johansen J, Johansen KM (2000) JIL-1, a chromosomal kinase implicated in regulation of chromatin structure, associates with the male specific lethal (MSL) dosage compensation complex. *J Cell Biol* 149:1005–1010
- Jones RS, Gelbart WM (1993) The *Drosophila* polycomb-group gene *Enhancer of zeste* contains a region with sequence similarity to *trithorax*. *Mol Cell Biol* 13:6357–6366
- Kelley RL, Kuroda MI (2000) The role of chromosomal RNAs in marking the X for dosage compensation. *Curr Opin Genet Dev* 10:555–561
- Lyman LM, Copps K, Rastelli L, Kelley RL, Kuroda MI (1997) *Drosophila* male-specific lethal-2 protein: Structure/function analysis and dependence on MSL-1 for chromosome association. *Genetics* 147:1743–1753
- Marín I, Baker BS (2000) Origin and evolution of the regulatory gene *male-specific lethal-3*. *Mol Biol Evol* 17:1240–1250
- Marín I, Franke A, Bashaw GJ, Baker BS (1996) The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. *Nature* 383:160–163
- Marín I, Siegal M, Baker BS (2000) The evolution of dosage compensation mechanisms. *Bioessays* 22:1106–1114
- Meyer BJ (2000) Sex in the worm. Counting and compensating X-chromosome dose. *Trends Genet* 16:247–253
- Nagase T, Kikuno R, Nakayama M, Hirokawa M, Ohara O (2000) Prediction of the coding sequences of unidentified human genes. XVIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro. *DNA Res* 7:273–281
- Neal KC, Pannuti A, Smith ER, Lucchesi JC (2000) A new human member of the MYST family of histone acetyl transferases with high sequence similarity to *Drosophila MOF*. *Biochim Biophys Acta* 1490:170–174
- Nicholas KB, Nicholas HB Jr (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the authors ([www.cris.com/~ketchup/genedoc.shtml](http://www.cris.com/~ketchup/genedoc.shtml))
- Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *CABIOS* 12:357–358
- Palmer MJ, Mergner VA, Richman R, Manning JE, Kuroda MI, Lucchesi JC (1993) The *male-specific lethal-one (msl-1)* gene of *Drosophila melanogaster* encodes a novel protein that associates with the X chromosome in males. *Genetics* 134:545–557
- Pannuti A, Lucchesi JC (2000) Recycling to remodel: Evolution of dosage-compensation complexes. *Curr Opin Genet Dev* 10:644–650
- Pardo PS, Leung JK, Lucchesi JC, Pereira-Smith OM (2002) MRG15, a novel chromodomain protein, is present in two distinct multiprotein complexes involved in transcriptional activation. *J Biol Chem* 277:50860–50866
- Ruiz MF, Esteban MR, Donoro C, Goday C, Sanchez L (2000) Evolution of dosage compensation in Diptera: The gene *maleless* implements dosage compensation in *Drosophila* (Brachycera suborder) but its homolog in *Sciara* (Nematocera suborder) appears to play no role in dosage compensation. *Genetics* 156:1853–1865
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sanjuán R, Marín I (2001) Tracing the origin of the compensasome: Evolutionary history of DEAH helicase and MYST acetyltransferase gene families. *Mol Biol Evol* 18:330–343
- Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC (2000) The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol Cell Biol* 20:312–318

- Song JY, Leung T, Ehler LK, Wang C, Liu Z (2000) Regulation of meristem organization and cell division by *TSO1*, and *Arabidopsis* gene with cysteine-rich repeats. *Development* 127:2207–2217
- Tatusova TA, Madden TL (1999) Blast 2 sequences—A new tool for comparing protein and nucleotide sequences. *FEMS Microbiol Lett* 174:247–250
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: Flexible strategies for multiple sequence alignments aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- Turner BM, Birley AJ, Lavender J (1992) Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. *Cell* 69:375–384
- Zhou S, Yang Y, Scott MJ, Pannuti A, Fehr KC, Eisen A, Koonin EV, Fouts DL, Wrightsman R, Manning JE, Lucchesi JC (1995) *Male-specific lethal 2*, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO J* 14:2884–2895