Tracing the Origin of the Compensasome: Evolutionary History of DEAH Helicase and MYST Acetyltransferase Gene Families

Rafael Sanjuán and Ignacio Marín

Instituto Cavanilles de Biodiversidad y Biología Evolutiva and Departamento de Genética, Universidad de Valencia, Spain

Dosage compensation in Drosophila is mediated by a complex of proteins and RNAs called the "compensasome." Two of the genes that encode proteins of the complex, *maleless* (*mle*) and *males-absent-on-the-first* (*mof*), respectively, belong to the DEAH helicase and MYST acetyltransferase gene families. We performed comprehensive phylogenetic and structural analyses to determine the evolutionary histories of these two gene families and thus to better understand the origin of the compensasome. All of the members of the DEAH and MYST families of the completely sequenced *Saccharomyces cerevisiae* and *Caenorhabditis elegans* genomes, as well as those so far (June 2000) found in *Drosophila melanogaster* (for which the euchromatic part of the genome has also been fully sequenced) and *Homo sapiens,* were analyzed. We describe a total of 39 DEAH helicases in these four species. Almost all of them can be grouped in just three main branches. The first branch includes the yeast PRP2, PRP16, PRP22, and PRP43 splicing factors and their orthologs in animal species. Each PRP gene has a single ortholog in metazoans. The second branch includes just four genes, found in yeast (*Ecm16*) and Drosophila (*kurz*) and their orthologs in humans and Caenorhabditis. The third branch includes (1) a single yeast gene (YLR419w); (2) six Drosophila genes, including *maleless* and *spindle-E/homeless*; (3) four human genes, among them the ortholog of *maleless,* which encodes RNA helicase A; and (4) three *C. elegans* genes, including orthologs of *maleless* and *spindle-E.* Thus, this branch has largely expanded in metazoans. We also show that, for the whole DEAH family, only MLE and its metazoan orthologs have acquired new protein domains since the fungi/animals split. We found a total of 17 MYST family proteins in the four analyzed species. We determined putative orthologs of *mof* in both *C. elegans* and *H. sapiens,* and we show that the most likely ortholog in yeast is the *Sas2* gene. Moreover, a paralog of *mof* exists in Drosophila. All of these results, together with those found for a third member of the compensasome, *msl-3,* suggest that this complex emerged after the fungi/animals split and that it may be present in mammalian species. Both gene duplication and the acquisition of new protein modules may have played important roles in the origin of the compensasome.

Introduction

Dosage compensation in drosophilid flies is achieved by an increase in transcriptional activity of the single male X chromosome. It is mediated by the modification of the chromatin structure of that chromosome by a complex formed by at least five proteins, encoded by the genes known as *male-specific lethal*s (*msl*s) (the five genes are called *msl-1, msl-2, msl-3, maleless* [*mle*], and *males-absent-on-the-first* [*mof*], respectively), and two noncoding RNAs, encoded by the *roX* genes (reviewed in Baker, Gorman, and Marín 1994; Lucchesi 1998; Franke and Baker 2000; Kelley and Kuroda 2000). This complex has been called the "compensasome" (see Franke and Baker 1999). In a wild-type male, compensasomes bind to hundreds of specific sites along the X chromosome. It has been suggested that binding starts in a few X-chromosome sites, including the *roX* genes, from which compensasome assembly and spreading would occur (Kelley et al. 1999; Meller et al. 2000). An understanding of what regions of the MSL proteins are involved in the assembly and/or biochemical action of the complex is emerging. All compensasome proteins have characteristic domains. Thus, MSL-1 has an acidic domain (Palmer et al. 1993), and it has been shown to contain two other regions important for

compensasome assembly, including a leucine-zipper-like coiled-coil domain (Scott et al. 2000). MSL-2 has a ring finger domain and another coiled-coil domain, both involved in protein-protein interactions (Bashaw and Baker 1995; Kelley et al. 1995; Zhou et al. 1995; Copps et al. 1998; Scott et al. 2000). Finally, MSL-3 has a chromodomain and a third coiled-coil domain, which are also likely to act to stabilize the compensasome (Koonin, Zhou, and Lucchesi 1995; Marín and Baker 2000). *msl-3* belongs to a recently described family of genes of unknown function (Marín and Baker 2000). In contrast, MLE and MOF belong to families for which extensive functional information is available. MLE is a member of the DEAH family of helicases (Kuroda et al. 1991; reviewed in de la Cruz, Kressler, and Linder 1999), and it has been shown that purified MLE has ATPase as well as RNA and DNA helicase activities (Lee et al. 1997). MOF is a member of the MYST family of acetyltransferases (Hilfiker et al. 1997). Several genes of this family, including *mof* itself, have been shown to encode proteins with histone acetyltransferase activity (Borrow et al. 1996; Reifsnyder et al. 1996; Smith et al. 1998; Champagne et al. 1999; Iizuka and Stillman 1999; Akhtar and Becker 2000). Interestingly, MOF, apart from an acetyltransferase domain, also contains a chromodomain quite similar to the one found in MSL-3, as well as a C_2HC zinc finger domain (Hilfiker et al. 1997).

In spite of all this information, the precise biochemical function of the complex is still incompletely understood. It is known that the male X chromosome is enriched in a particular isoform of histone H4, acetylat-

ed at lysine 16 (Turner, Birley, and Lavender 1992). This enrichment depends on the presence of compensasomes (Bone et al. 1994). The recent demonstrations that a partially purified complex containing the five MSL proteins and roX2 RNA (Smith et al. 2000), or simply MOF alone (Akhtar and Becker 2000), especially acetylate histone H4 at lysine 16, strongly suggest that MOF is directly responsible for the male- and X-chromosome-specific histone acetylation pattern. However, whether the other proteins or RNAs of the compensasome play a role apart from contributing to MOF deployment on the X chromosome is unknown. MLE helicase activity is obviously a good candidate for a second, independent, biochemical action of the complex. Meller et al. (2000) provided evidence suggesting that MLE is required for *roX2* RNA incorporation into the compensasome.

The evolution of dosage compensation mechanisms is still poorly understood (see reviews by Marín, Siegal, and Baker 2000; Meller 2000). This work tries to shed some light on the evolutionary origin of the compensasome. It has been determined that the compensasome is at least 60 Myr old by showing that the male- and chromosome-specific binding of several of the MSL proteins is conserved in drosophilid species of genera other than Drosophila (Marín et al. 1996). Beyond this phylogenetic range, the closest data come from nematodes, which diverged from insects about 550 MYA (Conway Morris 1998). It is known that the nematode *Caenorhabditis elegans* uses a totally different biochemical mechanism to achieve dosage compensation (based on chromatin condensation) that requires a set of proteins unrelated to those of the compensasome (see reviews by Cline and Meyer 1996; Lucchesi 1998; Meyer 2000). This observation notwithstanding, the compensasome may be older than the nematode-arthropod dichotomy. First, it is possible that dosage compensation was mediated by a Drosophila-type mechanism before the arthropod/nematode split, and the mechanism now found in *C. elegans* was developed secondarily after that split (see Marín, Siegal, and Baker 2000). Second, the compensasome might be involved in different organisms in independent processes requiring chromatin activation, having been coopted for dosage compensation only in particular species.

If the compensasome is indeed an ancient chromatin-modifier complex, we expect to find orthologs of the Drosophila compensasome genes in other species. So far, and we confirmed this when preparing this study, genes obviously related to Drosophila *msl-1* or *msl-2* have not been found. Although such a negative result does not demonstrate that *msl-1* and *msl-2* have no relatives in other species (the alternative being that they are too divergent to be detected as such), it precludes further studies. In contrast, genes related to *mle, mof,* and *msl-3* have been found in yeasts, nematodes, and mammals (Kuroda et al. 1991; Hilfiker et al. 1997; Marín and Baker 2000; Neal et al. 2000). The most complete evolutionary study available so far for a member of the compensasome is the one performed for *msl-3* and its relatives. Marín and Baker (2000) have shown that *msl-3* is part of a small family of genes that has a single member in yeasts but two paralogous genes (*msl-3* and *mrg15*) in both humans and Drosophila. One of these duplicated genes, most likely *msl-3* itself, has been lost in *C. elegans.* These results suggest that the *msl-3/mrg15* duplication occurred after the fungi/animals split but before the different metazoan lineages diverged, that is, at least 600 MYA (see Conway Morris 1998). For *mle* and *mof,* the precise phylogenetic relationships with other genes of their respective protein families have never been determined.

We performed exhaustive database searches for genes related to *mle* and *mof* in four different species—the completely sequenced yeast *Saccharomyces cerevisiae* and the nematode *C. elegans,* and two species whose genome projects are well advanced, *Drosophila melanogaster,* for which the euchromatic portion of the genome has been completed, and *Homo sapiens*—in order to gain insight into the origin and evolution of these two genes. We combined phylogenetic sequence analysis and structural comparisons to determine the relationships among the genes. Our results suggest that *mle* and *mof,* as has also been shown for *msl-3* (Marín and Baker 2000), derive from duplications that occurred at some point after the fungi/animals split. We also show that the MLE protein and its orthologs in other animal species are the only members of the DEAH family that have acquired new protein domains since the fungi/animals divergence. We discuss the implications of these findings with respect to the origin of the compensasome.

Materials and Methods

There is some confusion in the literature with respect to what proteins belong to the DEAH family of helicases. We followed the rigorous definition of de la Cruz, Kressler, and Linder (1999), which is based on sequence similarity, not just the presence or absence of a DEAH motif (see also Jankowsky and Jankowsky 2000).

Data Mining

All database searches were performed online and finished in early June 2000. The databases analyzed were (1) the nonredundant, month, and dBest databases at the National Center for Biotechnology Information (NCBI) web pages (http://www.ncbi.nlm.nih.gov/); (2) for *D. melanogaster* sequences, the databases of the European *Drosophila* Genome Project (EDGP) and the Berkeley *Drosophila* Genome Project (BDGP) web sites (http://edgp.ebi.ac.uk and http://www.fruitfly.org, respectively); (3) for *C. elegans,* the database at the Sanger Center web pages (http://www.sanger.ac.uk). In order to detect all the available sequences corresponding to DEAH or MYST proteins, database searches were performed iteratively, using first the sequences of the MLE and MOF proteins, and then the sequences with the highest similarity to those two. At the NCBI web site, comparisons were performed using the TBLASTN and PSI-BLAST programs (Altschul et al. 1997). At the EDGP, BDGP, and Sanger Center web sites, the available TBLASTN program was used.

Sequence Comparisons

Once all the sequences corresponding to DEAH or MYST proteins were detected (i.e., when the searches became saturated), the regions with the highest similarity were aligned using the multiple-alignment program ClustalX (Thompson et al. 1997). These regions are the helicase core domain of DEAH helicases (see de la Cruz, Kressler, and Linder 1999; corresponds to amino acids 384–873 in MLE) and the zinc finger and acetyl-CoA-binding domains and other highly conserved regions described by Reifsnyder et al. (1996) and Hilfiker et al. (1997) for MYST acetyltransferases. We called these conserved regions ''MYST core domain.'' We were able to extend the alignment for some proteins of the DEAH family (i.e., yeast PRP splicing factors) up to about 200 amino acids C-terminal of the helicase core domain. However, because similarity was low for other members of the DEAH family, that region was not included in the analyses presented below. In some cases, the ORF Finder program, also available at the NCBI web site, was used to establish ORFs. Ambiguous regions that may or may not correspond to introns in genes of the DEAH family were kept in the sequences. This did not affect the conclusions presented below, because positions with gaps in one or more sequences were eliminated to obtain the DEAH family phylogenetic tree.

Once the output of the ClustalX program was obtained, it was refined manually by editing the sequences with the program GeneDoc (Nicholas and Nicholas 1997). Phylogenetic trees were obtained using the neighbor-joining method (Saitou and Nei 1987), and we used bootstrap analysis to determine the reliability of the topology obtained, performing a total of 1,000 replicates. The neighbor-joining and bootstrap algorithms used were also those included in ClustalX. The program TreeView, version 1.5 (Page 1996), was used to draw the phylogenetic trees. GeneDoc was also used to highlight the similarities among multiple aligned sequences.

Structural Analyses

To search for other significant domains in these proteins, we analyzed the sequences of the regions previously excluded from the multiple alignments. Each one of those sequences was compared with the sequences in the databases described above using BLASTP or TBLASTN, in order to detect similarities corresponding to domains characteristic of just one or a few of the proteins of these families. The procedures of multiple alignment, generation of phylogenetic trees, and determination of bootstrap values described in the previous section were also used to obtain the chromodomain and PHD finger results shown in figures 5 and 6.

Results

Evolutionary History of the DEAH Family

A total of 39 DEAH proteins were found in the four eukaryotic genomes analyzed. Figure 1 shows the phylogenetic tree derived from the alignment of those proteins (available at the European Bioinformatics Institute;

see http://www3.ebi.ac.uk/Services/align/listali.html; accession number DS44798). It also includes the bootstrap values that support some of the deepest branches, which correspond to orthology groups. Bootstrap data shown in figure 1 strongly support three main groups of DEAH proteins in eukaryotes. There are two large groups with many members, as well as a third, small, group formed by just four proteins. One of the large groups includes the four yeast DEAH proteins known to be involved in splicing (PRP2, PRP16, PRP22, and PRP43; see de la Cruz, Kressler, and Linder [1999] for a review) and their putative orthologs in animals. For each of these four yeast genes, we detected a single ortholog in each metazoan species. It is very likely that these genes are true orthologs and that they have the same role in all these species. In several cases, direct data support similar roles of these proteins in yeasts and animals (Ono, Ohno, and Shimura 1994; Gee et al. 1997; Zhou and Reed 1998). The apparent exception is the *C. elegans mog* genes (*mog-1, mog-4,* and *mog-5*; Puoti and Kimble 1999, 2000), known to be involved in sex determination and, in the case of *mog1,* without an apparent role in splicing (Puoti and Kimble 1999). These particular cases will be discussed below.

In the second, small, group formed by *S. cerevisiae Ecm16* and its relatives, the available data again suggest that only a single gene is present in each species. In contrast, in the third, large, group of proteins, the one that includes MLE, there is just one yeast protein (YLR419w), but there are multiple proteins in metazoans: six in Drosophila, three in nematodes, and four in humans (excluding the sequence we called ''Hs 13,'' which corresponds to a pseudogene derived from the *RNA helicase A* gene). Only four proteins remain of uncertain classification, namely, yeast JA2 and the products of two still-undescribed Drosophila genes (CG3225 and CG4901) and one *C. elegans* gene (T05E8.3). In our tree (fig. 1), they appear together in an intermediate position, so it is possible that they form a fourth monophyletic branch.

In two cases, we observed the emergence of new protein domains in particular groups of proteins of the DEAH family. First, yeast PRP22 and its metazoan orthologs have an S1 domain, characteristic of many RNA-binding proteins (Gribskov 1992; Bycroft et al. 1997). Second, MLE and its closest relatives (all in metazoan species) have two characteristic dsRNA-binding domains (as first found by Gibson and Thompson 1994), as well as a C-terminal GGY repetitious motif (described by Kuroda et al. 1991). Thus, these are the only acquisitions of new domains by proteins of this family since the fungi/animals split.

The MYST Family of Acetyltransferases

Figures 2 and 3, respectively, show the alignment of the most conserved region of this family of proteins, which we called the MYST core domain (see *Materials and Methods*), and the corresponding phylogenetic tree. In searching for additional protein domains encoded by these genes, we found three of them (fig. 4). First, a

Dm SpindleE 875
Ce F52B5.3
Sc YLR419w
963
Dm CG8915
Dm CG1582
Dm CG9323
1000
Hs KIAA1488
Hs KIAA0890
Ce T07D4.3 1000
Ce Y37E11AM.c
Dm CG12211
Dm Maleless
Hs HelicaseA
Hs 13
1000
Hs KIAA1517
Dm Kurz
Ce C06E1.10
Sc ECM16
dsRNA-binding domains, GGY repeats
Hs ALO79292.1
Dm CG3225
Ce T05E8.3
S1 domain
Dm CG4901
1000
Dm CG8241
Hs HRH1
Ce MOG5
Sc PRP22
Sc JA2
Sc PRP16
Ce MOG1
Hs PRP16
Dm CG1375/1405
986
948
Ce MOG4
Dm CG10689
Hs HEL#3
Sc PRP2
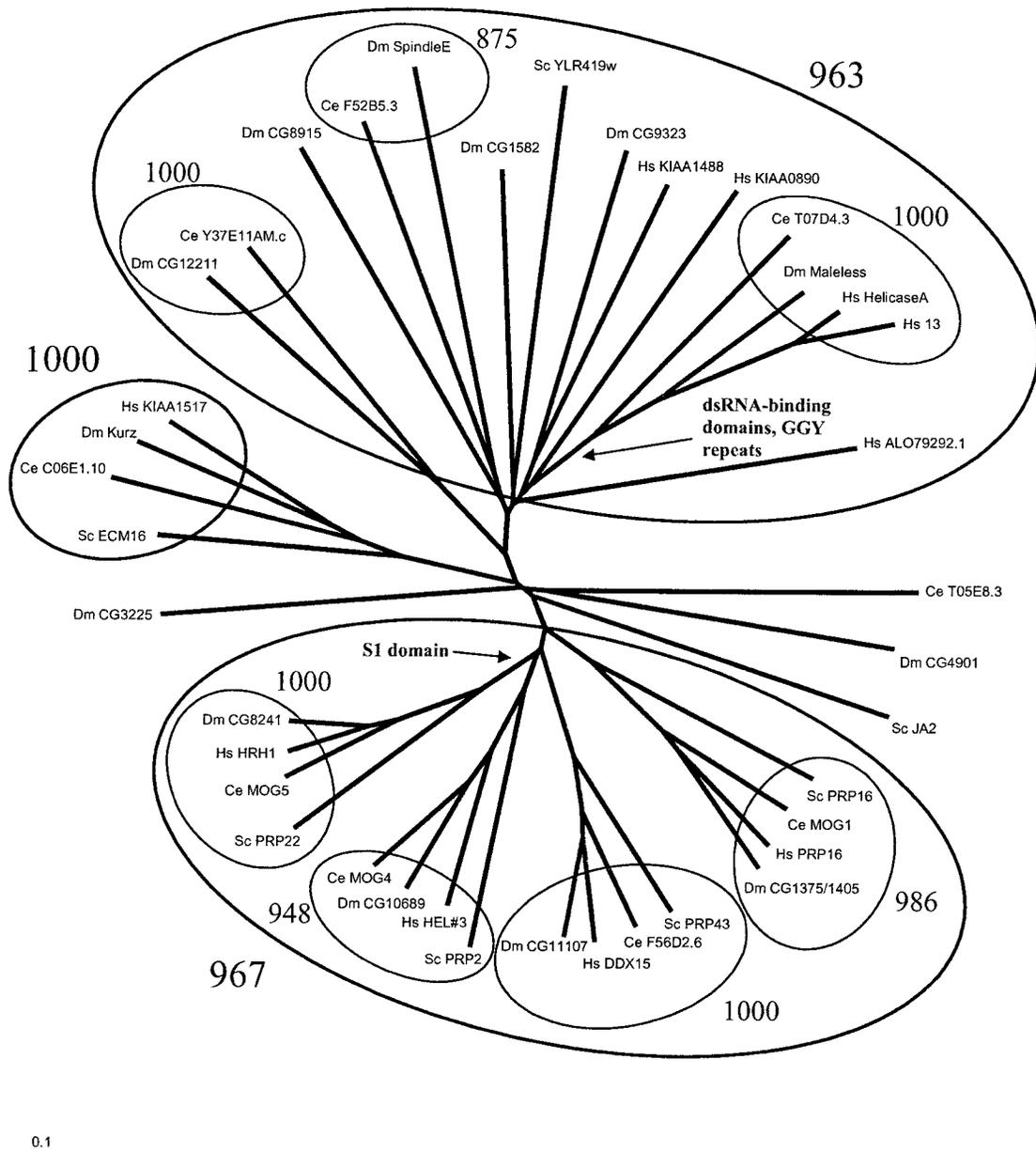Dm CG11107
Sc PRP43
Ce F56D2.6
Hs DDX15
967
1000
0.1

FIG. 1.—Phylogenetic tree for DEAH helicases. Numbers refer to the bootstrap values. The most significant associations have been highlighted, including all the related proteins (in ellipsoidal boxes). Bootstrap values less than 600 or those connecting very similar proteins within an orthology group (e.g., within the PRP groups) are, for simplification, not shown. The branches of the three in which new functional domains emerged are also detailed. Dm CG1375/1405 refers to the fact that this gene had been so far mistakenly described as two different genes.

chromodomain was found in seven proteins, namely, *S. cerevisiae* ESA1, *D. melanogaster* MOF, *H. sapiens* TIP60 and MOF (assembled by combining data from overlapping human clones with accession numbers AI623700.1, AI934823.1, AI583211, AL50395.1, and AW131094.1; after we obtained this sequence, we found that it corresponded to the gene very recently described by Neal et al. [2000]), and the previously undescribed EG0007.7 (*D. melanogaster*), VC5.4, and K03D10.3 (*C. elegans*). A second domain found was the PHD finger, which was present in four sequences (*C. elegans* gene encoding C34B7.1, Drosophila CG11290, and human genes encoding MOZ and MORF). Finally, *H. sapiens*

HBO1 and *D. melanogaster* CG5229 contained an additional zinc finger domain of the C_2HC type. Two other characteristic regions, one of them acidic and the other rich in proline and glutamine, were found in human MOZ and MORF and in the *C. elegans* protein R07B5.8. The *S. cerevisiae* protein SAS3 also had an acidic region in its C-terminus (fig. 4).

The combined phylogenetic and structural analyses suggest that there are three groups of MYST proteins. The first group would contain the relatives of the human proteins MOZ and MORF. Five of the members of this group (*H. sapiens* MOZ, MORF, and HBO1 and *D. melanogaster* CG5229 and CG11290) show substantial

```
                              *        20        *        40        *        60
Ce K03D10.3  : ITI AHE-  LA  Y  --F PDCENLD-------------------IY  Y  L TPH :  37
Sc SAS2      : IQF LNKKFST  G AVYFDPETKRLGCSETKGQLSSVSNSQYWLDTLF  Y FK TDD :  60
Hs MOF       : IHI NYE-  DA  F  --F EDYGKQPK-------------------LW  Y  K MKY :  38
Dm MOF       : LQF NYE- DT  F  --F EEYGKART-------------------LY  Y  K MRF :  38
Dm CG1894    : VQF RYE- ETTSS  --Y VINDKATT-------------------IY  F  K MCL :  38
Sc ESA1      : IIM KYE- EP  F  --Y IELTDEDF-------------------IY DDF  Q FGS :  38
Ce VC5.4     : IEL RSR- QP  PA  --Y QQLTSLDC-------------------IY  F  K LKS :  38
Hs TIP60     : IEL RHR- KP  F  --Y QELTTLPV-------------------LY  F  K GRS :  38
Dm EG0007.7  : IEL RHR- KP  F  --Y QELCQMPC-------------------IY  F  K RKS :  38
Hs HBO1      : IAF RYE- DT  H  --Y EEYARLGR-------------------LY  F  K MKS :  38
Dm CG5229    : ISM KYK- KV  Q  --Y DDAARLPK-------------------MY  F  R QKS :  38
Dm CG11290   : IQI KWD- ET  S  --F QEYARLLK-------------------LF  F  K TKS :  38
Hs MOZ       : IEF KYE- HT  S  --Y QEYSRLPK-------------------LY  F  K MKS :  38
Hs MORF      : IEF KYE- QT  S  --Y QEYARLPK-------------------LY  F  K MKS :  38
Ce R07B5.8   : IHF AFI- KT  G  --F AEFINVKK-------------------LF  F FF ARS :  38
Sc SAS3      : IVLRNYE- KP  T  --F EHINQNKM-------------------VF  F  K MTS :  38
Ce C34B7.1   : LHF TGKACKAI S A--YEEPLHSSAH-------------------LYA KF  HTTNL :  39

                              *        80        *       100        *       120
Ce K03D10.3  : HERFKQ IDT KVRQ- P N  Y KDH  V  V GSGQKLY  CLCL SKLFM HK L  :  96
Sc SAS2      : QTRFVG VAS PFQYRVP KIK  SPEY IRRVKGSKYQLF  CLCLFTKLYL NK M  : 120
Hs MOF       : EKSYRF LGQ QWRQ- P K  Y KSN  V  V GKDHKIY  NLCL AKLFL HK L  :  97
Dm MOF       : RSSYAY LHE DRRR- P R  Y KGN  I  V GKEESLY  LLCL AKLFL HKVL  :  97
Dm CG1894    : RKSYSY LYD CKKRC- P SL  Y KDN YI  V GHKEQLY  CLCL SKLFLENKKIL :  97
Sc ESA1      : KKQYERYRKK TLRH- P N  Y DDY  F  I GRKQRTW RNLCL SKLFL HK L  :  97
Ce VC5.4     : KTCLKR MXK AMCH- P N  YSHDK  F  I GRKNKSYA NLCL AKLFL HK L  :  97
Hs TIP60     : LKCLQR LTK DLRH- P N  Y KGT  F  I GRKNKSYS NLCL AKCFL HK L  :  97
Dm EG0007.7  : RKCLER LSK NLRH- P N  Y KHT  F  I GRKNKVYA NLCL AKLFL HK L  :  97
Hs HBO1      : QTILRR MAK CVWKH- P D  Y KGS  V  V GKKNKIY  NLCL AKLFL HK L  :  97
Dm CG5229    : ETGIKR AEK CVWRH- P D  Y KGK QV  V GKRYKQY  HLCL AKFFL HK L  :  97
Dm CG11290   : RSVLDR QNK CIWKQ- P T  F QGN  V  V GNVNKIY  NLCL AKFFL HK L  :  97
Hs MOZ       : RTILQQ MKK CGWFH- PAN  Y KNN  V  V GNVSTIY  NLCL AKLFL HK L  :  97
Hs MORF      : KNILLR SKK CGWFH- PAN  Y RKD  V  V GNMSKIY  NLCL AKLFL HK L  :  97
Ce R07B5.8   : DEIMQN AKK CMLRA- P L  Y KGD  V  V GRLQKEY  TLCL SRMFLESK V  :  97
Sc SAS3      : RYTFYR QLK CLTFK- P N  Y DGK  V  I GRENVLY  NLCL AKCFI SK L  :  97
Ce C34B7.1   : KEDLIV WDI CNAKH- P T  Y NDG AF  V GAVQKKY  DLCL SKLFIASK L D :  98

                              *       140        *       160        *       180
Ce K03D10.3  : VDDF FY CET EH --------A I GY R VESA--NNLACIMVFPP  KG  : 145
Sc SAS2      : KVDH FY VYET-GSTK---------P GF KDLVSYQQNNLACILIFPP  RGL  : 169
Hs MOF       : VEPF FY TEV RQ --------A I GY K ESPDGNNVACILT PP  RG  : 148
Dm MOF       : MDPF FY CET KE --------S I GY K KSLENYNVACILV PPH  KG  : 148
Dm CG1894    : SSSS FY CLK KD --------E FAGY AR TML-NINLNCILV PP M KG  : 147
Sc ESA1      : VDPF FYC TRR EL --------H I GY K ESADGYNVACILT PQ  MG  : 148
Ce VC5.4     : TDPF FY TEE EK --------H I GY K ESAEEYNVACILV PP  KG  : 148
Hs TIP60     : TDPF FY TEY CK --------F I GY K ESTEDYNVACILT PP  RG R : 148
Dm EG0007.7  : TDPF FY TEF SR --------F I GY K ESTEDYNVACILT PP  KG  : 148
Hs HBO1      : VEPF FY TEA NT --------C L GY K NSFLNYNVSCIIT PQ M QG  : 148
Dm CG5229    : VEPF FY TLA VD --------C I GY K NSF-YNVSCIIT PP  KG  : 146
Dm CG11290   : VEPF FY TKN QS --------C L GY K HCTQKYNVSCIIT PQ  QG  : 148
Hs MOZ       : VEPF FY TQN VK --------C I GY K HCQQKYNVSCIMI PQ  KG  : 148
Hs MORF      : VEPF FY TKN EK --------C I GYL K LCQQKYNVSCIMI PQH  QG  : 148
Ce R07B5.8   : TEPF FY TIN DI --------C FAGY K YEPDVNNLSCIMT PC  EMG  : 148
Sc SAS3      : VEPF FY TERE TENHPYQNAAKF GY K FNSNDYNLSCILT PI  KG  : 157
Ce C34B7.1   : EVET FY CEI-TTE --------YVI GY K NPSKNNNLSCLLV PMV  MG  : 149
```
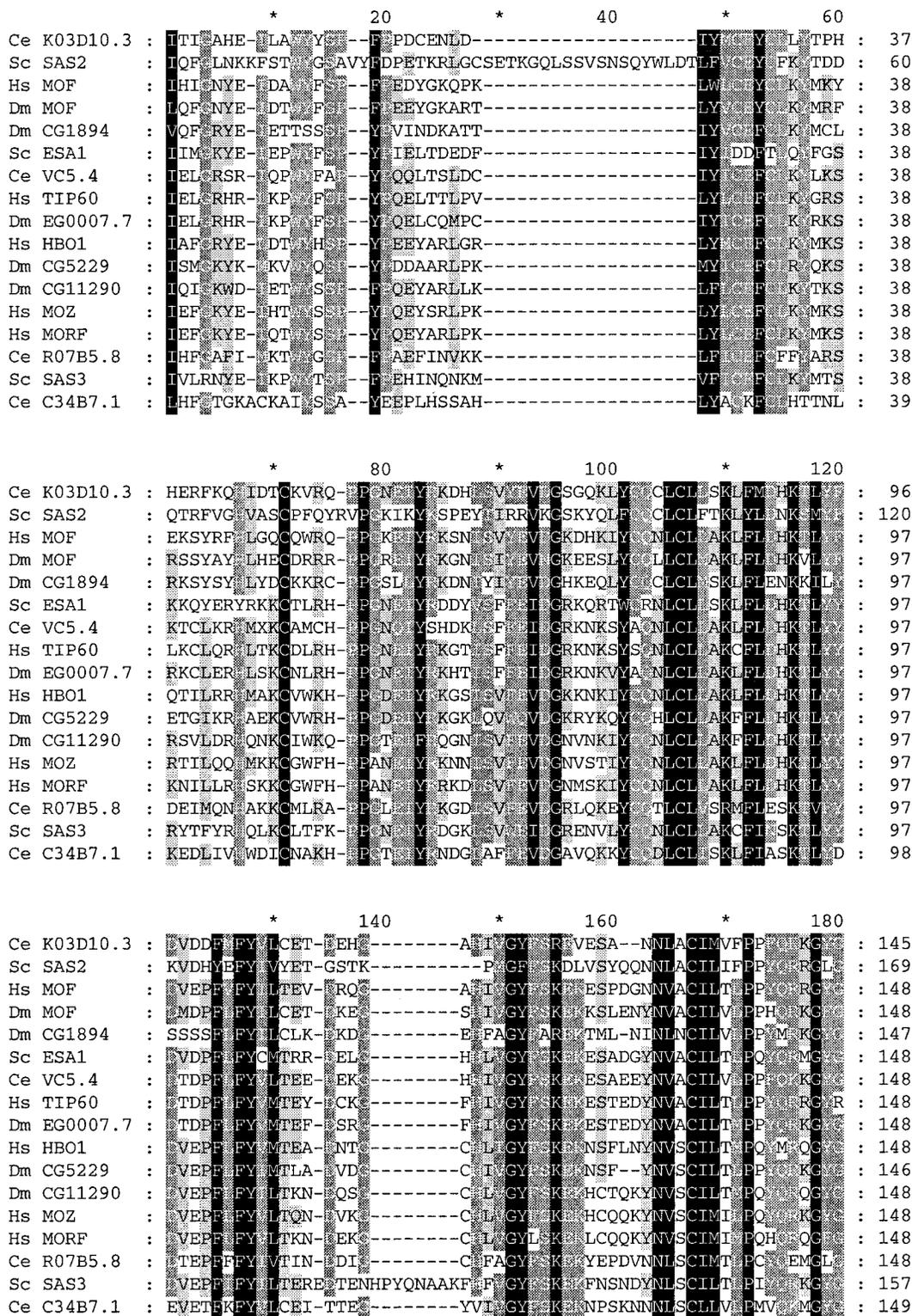
FIG. 2.—Multiple alignments of the MYST core domain. Sequences are ordered according to figure 3, clockwise from Ce K03D10.3.

similarities in their MYST core domain sequences (fig. 3). The other three (*S. cerevisiae* SAS3 and *C. elegans* C34B7.1 and R07B5.8), although with less-related MYST core domain sequences, have significant structural similarities with MOF and MORF (fig. 4). A second group would include a single protein in yeasts (ESA1) and metazoans (human TIP60, *D. melanogaster* EG0007.7, and *C. elegans* VC5.4). The monophyly of this group is supported by both strong sequence similarity (fig. 3) and structural data, because all of them contain a chromodomain (fig. 4). Finally, a third group would be formed by *D. melanogaster* MOF and its clos-

```
                          *       200       *       220       *       240
Ce K03D10.3 : LLIQ SYELS REGYI MP  PLS   K SYRSYWWWR  KLFHIHQ------------ : 193
Sc SAS2     : LLLIE SYKLSQLEGVISGP VPLSPF L GYLKYWSQI CWH IEGDLAH--------- : 220
Hs MOF      : FLIA SYELS LESTV SP  PLS   K SYRSYWSWV  EI RDFR------------ : 196
Dm MOF      : LLLJA SYELS KEGVI SP  PLS   R SYRSYWAYT  EL KTRCA----------- : 197
Dm CG1894   : LLIDLSYEIS REACI GPK PLSK AR CYLSYWGHI  NL RHHSS----------- : 196
Sc ESA1     : KLLIE SYELS KFNKV SP  PLS   L SYRAYWSDT  TL VEHQ------------ : 196
Ce VC5.4    : SILIE SYELS IFQKT SP  PLS   L SYRSYWSMA  KE FAFKRRHP-------- : 200
Hs TIP60    : LLIE SYELS VEGKT TP  PLS   L SYRSYWSQT  EI MGLK-SESGE------ : 201
Dm EG0007.7 : LLIE SYELS FEGKT SP  PLS   L SYRSYWAQT  EIFISQNPSTDGE------ : 202
Hs HBO1     : MLID SYLLS VEEKV SP  PLS  CL SYRSYWKEV  RY HNFQ------------ : 196
Dm CG5229   : LLID SYLLT VEGKI SP  PLS  CL SYRSYWKDV  DY CNRS------------ : 194
Dm CG11290  : FLID SYLLS BEGQL TP  PLS   R SYFSYWKSV  EY YKHRNYTK-------- : 200
Hs MOZ      : FLID SYLLS REGQA SP  PLS   R SYMAYWKSV  EC YHQNDKQ--------- : 199
Hs MORF     : FLID SYLLS REGQA SP  PLS   R SYLAYWKSV  EY YHHHERH--------- : 199
Ce R07B5.8  : FLIDISYALS KBKWF GP QPLSE  RKAYGGYWRTT ASC GRLKDELEFG------ : 202
Sc SAS3     : QFLME SYLLS KESKF TP  PLS   L TYRTFWKIKCAEV LKLRDSARRRSNNKNE : 217
Ce C34B7.1  : LLIDMSYELS LEMRI HP HPLS  CI AYRGYWRSS  CY REHRNYDR-------- : 201
```

```
                          *       260       *       280       *       300
Ce K03D10.3 : --GHTV ATFLSNESG AVDDI ST ITVRMCRQYKEPEF PGEW---Y RIHRKIVDHC : 248
Sc SAS2     : --YDKVT EDLSIVTG RVNDV LT KHLNCIGENNQIYLQ-------S N--SWLKLHG : 269
Hs MOF      : ---GTLS KDLSQMTS TQNDI ST QSLNMVKYWKGQHV -------C T-PKLVEEHL : 245
Dm MOF      : --PEQI KBLSEMSG THDDI YT QSMKMIKYWKGQNV -------C T-SKTIQDHL : 247
Dm CG1894   : --PDLV EBLSKATGFREEDI ST EFMGMTKYYKVDHI -------FYTTSSIIEDRR : 247
Sc ESA1     : ---KEI DELSSMTS TTTDI HTAKTINILRYYKGQHI -------F N-EDILDRYN : 245
Ce VC5.4    : --GEDI QDLSQSTS KREDV ST QQLDLYKYYKGSYI -------V S-DEKRQVYE : 250
Hs TIP60    : --RPQI NBLSEITS KKEDV ST QYLNLINYYKGQYI -------T S-EDIVDGHE : 251
Dm EG0007.7 : --KPTI SDLCECTS KKEDV ST QNLNLINYYKGQYI -------C N-RVIIEQHR : 252
Hs HBO1     : --GKEI KELSQETA NPVDI ST QALQMLKYWKGKHL --------KRQDLIDEWI : 246
Dm CG5229   : --GNTIA KDVSQVSL VGHDI ST QALGMMKYWKGKHI --------KKQDVLDEYE : 244
Dm CG11290  : -----I FKDIAIKTG AISDI ALAFELINFIKLRKNDGD RYQINV-K EWKKVLAHHN : 254
Hs MOZ      : -----IS KKLSKLTG CPQDITST HHLRMLDFRSDQFV --------RREKLIQDHM : 246
Hs MORF     : -----IS KALSRATG CPHDIATT QHLHMIDKRDGRFV --------RREKLILSHM : 246
Ce R07B5.8  : ---SGI KMIADDTG NCHDI EV CSLGWA------------------------ : 231
Sc SAS3     : DTFQQV NDIAKLTG IPTDV FG EQLQVLYRHKTRSLSSLDDFNYI KIDSWNRIEN : 277
Ce C34B7.1  : -----LN KDMCLATR APVDI NQ M-LDNMIVLKG--G -YSI---K GKRALKFPL- : 248
```

```
                          *       320
Ce K03D10.3 : VMCGYGKPPRLL DRTK R A - : 270
Sc SAS2     : TKRNWFK----- KDEY LIDD- : 286
Hs MOF      : KSAQYKK-PPIT DSVC K A - : 266
Dm MOF      : QLPQFKQ-PKLT DTDY V S - : 268
Dm CG1894   : GLAQFKK-PRLT HRNR S KT- : 268
Sc ESA1     : RLKAKKR---RT DPNR I K - : 264
Ce VC5.4    : KRIEAAK-KKTR NPAA Q R - : 271
Hs TIP60    : RAMLKRL---LR DSKC H T - : 270
Dm EG0007.7 : RAMDKRK---IR DSKC H T - : 271
Hs HBO1     : AKEAKRSNSNKT DPSC K T - : 268
Dm CG5229   : ERVKRRG-TFPK DDSC R Q - : 265
Dm CG11290  : KMANSKT--RII EPDC R S - : 274
Hs MOZ      : AKLQLNL-RPVD DPEC R T - : 267
Hs MORF     : EKLKTCS-RANE DPDS R T - : 267
Ce R07B5.8  : ---------------------- : -
Sc SAS3     : IYKTWSSKNYPR KYDK L E - : 299
Ce C34B7.1  : ----SQC-RRRF DPSQ F K - : 265
```

FIG. 2 *(Continued)*

est relatives (*S. cerevisiae* SAS2, *D. melanogaster* CG1894, *C. elegans* K03D10.3, and *H. sapiens* MOF), which also share both high sequence similarity and structural characteristics, with three of them also having a chromodomain (fig. 4). Both the yeast gene *Sas2* and Drosophila CG1894 have most likely secondarily lost the N-terminal chromodomain (fig. 4). To ascertain that all chromodomains and PHD finger domains detected had a monophyletic origin, we performed analyses comparing the MYST family proteins with other proteins that have those domains. The results are shown in figures 5 and 6. As expected, proteins of the MYST family
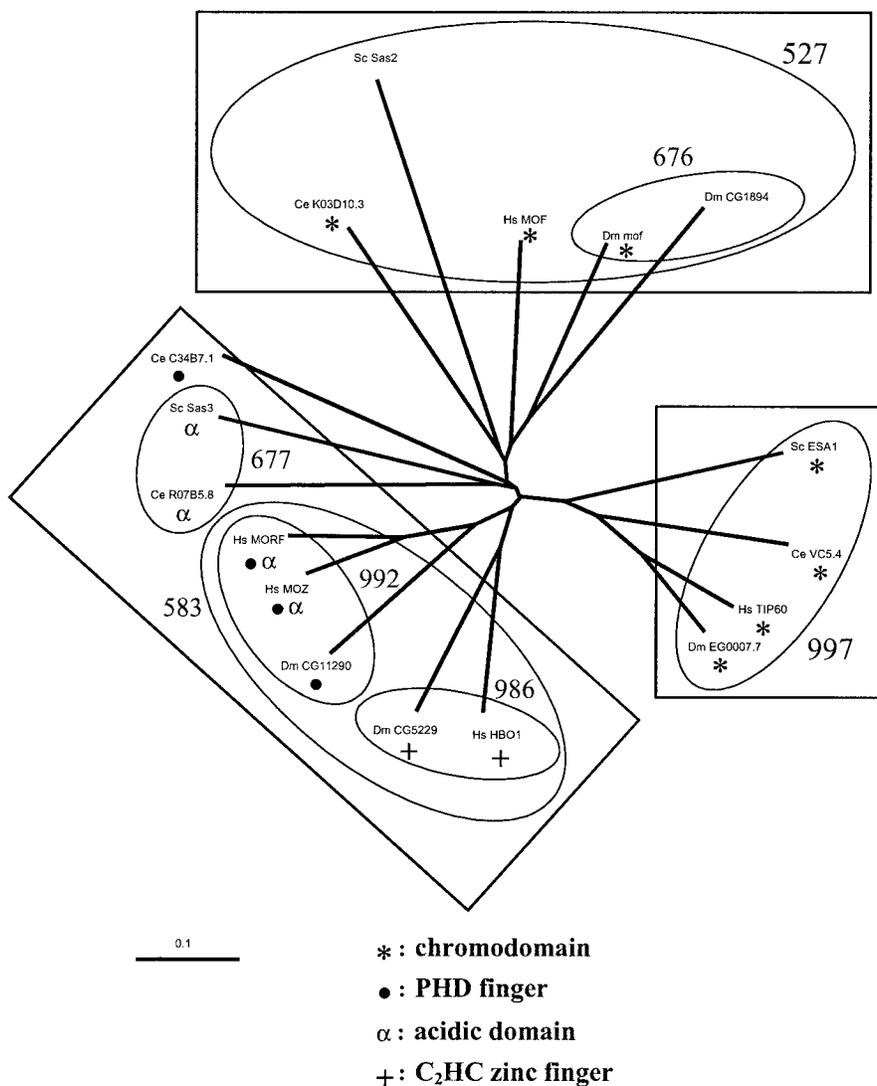
Fig. 3.—Phylogenetic tree for MYST acetyltransferases. Associations supported by bootstrap values greater than 500 are shown. The three main groups described in the text are also highlighted (in rectangular boxes). Domains detailed in figure 5 are also shown.

have very similar domains, a result that suggests a common origin. Interestingly, two proteins from amphibian species contain chromodomains that are very similar to those in proteins of the MYST family. Moreover, figure 5B shows that the chromodomains in MSL-3 and its relatives (Marín and Baker 2000) are much more similar to those in the MYST family proteins than the canonical chromodomains found in POLYCOMB or HP1 proteins. Figure 6 also shows that there is a close similarity among the PHD fingers of MYST family proteins to those in REQUIEM proteins (Gabig et al. 1994).

Discussion
Patterns of Diversification of DEAH Helicases in Metazoans

We determined the most likely orthologous genes in metazoan species for six of the seven yeast DEAH proteins (fig. 1). So far, five of those six proteins are represented by just one orthologous gene in metazoans. There was a seventh yeast gene, *Ja2,* for which the phy-

logenetic analysis yielded inconclusive results. Considering the topology shown in figure 1, it is reasonable to hypothesize that the other "orphan" genes left in our tree (*D. melanogaster* CG3225 and CG4901 and *C. elegans* T05E8.3) are orthologous to *Ja2.*

We conclude that the DEAH family of proteins has, in general, followed a conservative pattern of evolution. The apparent absence of gene duplications for five of the yeast genes in metazoans may correlate with most DEAH helicases retaining the same functions in fungi and animals. This has been already experimentally shown for three of the PRP human orthologs (Ono, Ohno, and Shimura 1994; Gee et al. 1997; Zhou and Reed 1998). However, recent studies that describe three *C. elegans* genes (*mog-1, mog-4,* and *mog-5*; Puoti and Kimble 1999, 2000) are in apparent conflict with this conclusion. According to data presented in Puoti and Kimble (2000) and the more extensive analysis shown in figure 1, these *mog* genes are the nematode orthologs of yeast *prp16, prp2,* and *prp22.* However, lack of *mog*
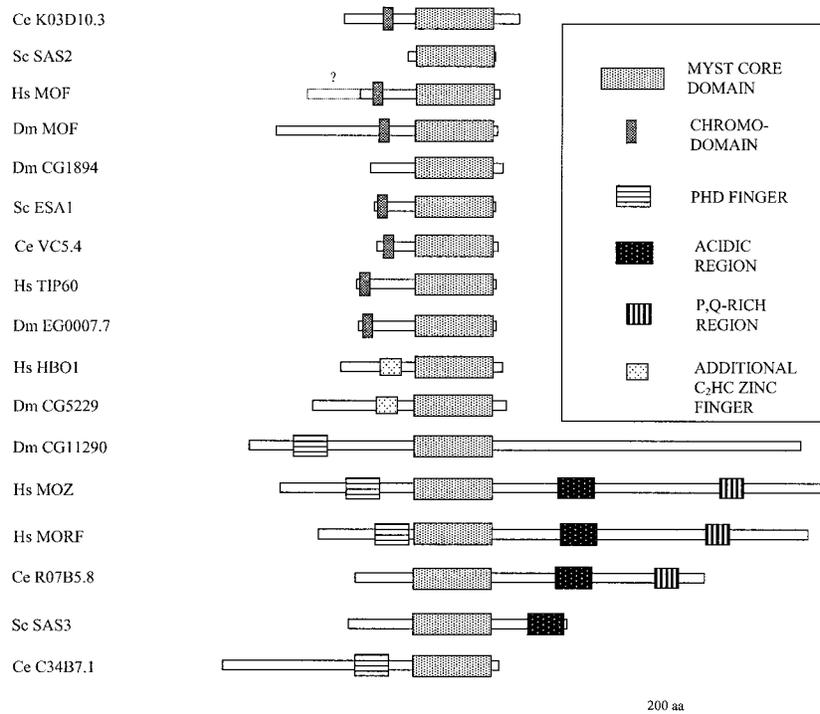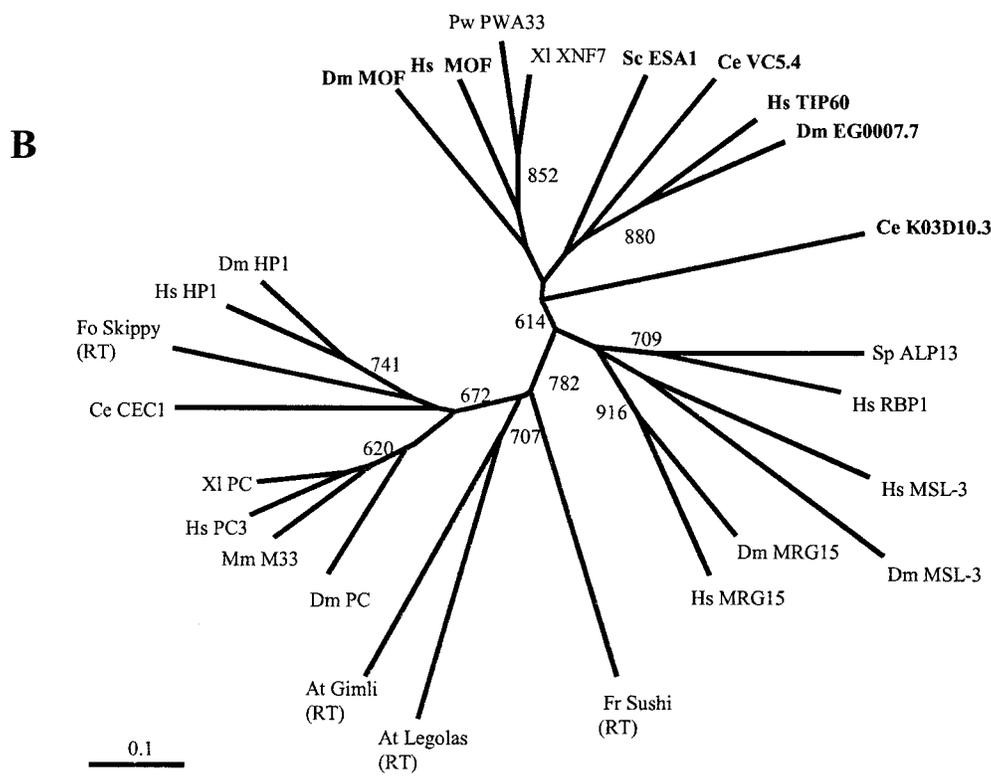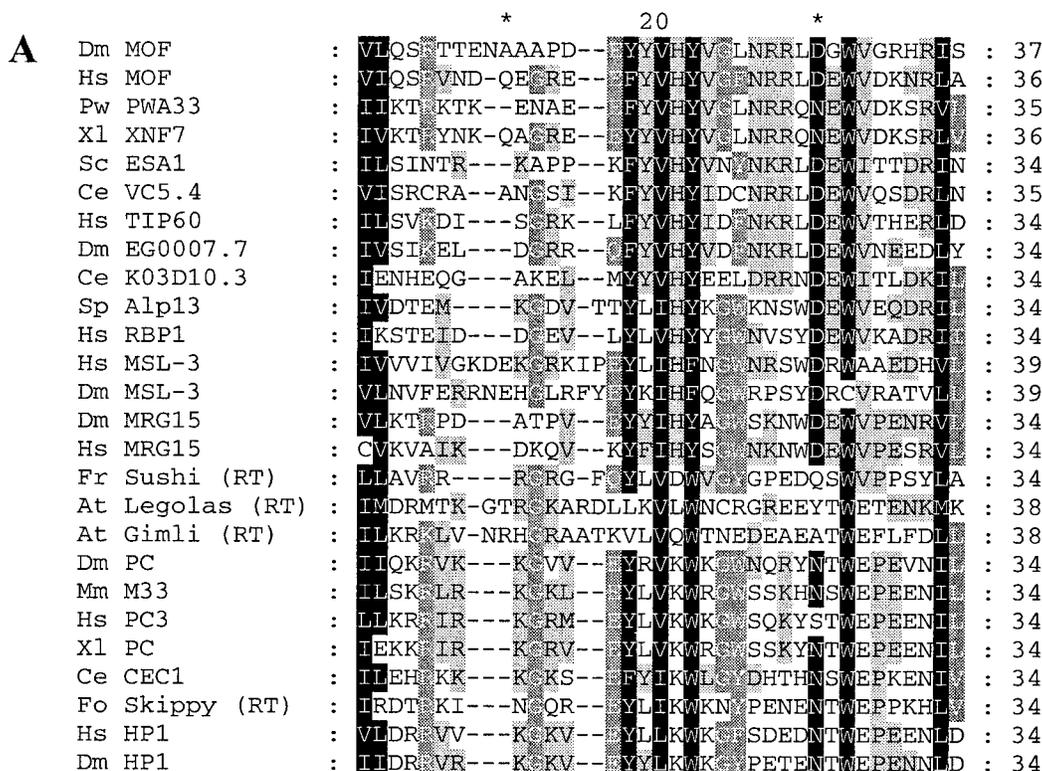
FIG. 4.—Structure of MYST acetyltransferases, showing the location of the MYST core domain plus the additional protein domains found (see text). The question mark in *Homo sapiens* MOF refers to the fact that the N-terminal end of this protein has not been characterized.

gene function blocks the transition from sperm production to oocyte production in *C. elegans* hermaphrodites, a phenotype characteristic of genes involved in sex determination in this species (Graham and Kimble 1993; Graham, Schedl, and Kimble 1993). Moreover, no apparent splicing alterations in *mog-1* mutants were found by Puoti and Kimble (1999). Therefore, a switch to a totally new function for these genes cannot be ruled out. However, a reasonable alternative is that these three *mog* genes indeed encode proteins involved solely in splicing and that the sex determination phenotypes of *mog* mutants are indirectly caused by abnormal splicing of some regulator of the target sex determination gene *fem-3* (as suggested by Puoti and Kimble 1999; see also Gallegos et al. 1998). In fact, worms lacking maternal *mog* products die early in embryogenesis (Graham and Kimble 1993; Graham, Schedl, and Kimble 1993). It is thus possible that the lack of splicing alterations found in *mog-1* mutants by Puoti and Kimble (1999) is due to the presence in those mutants of sufficient maternal protein to allow (quasi-)normal splicing, with only a few genes (including a *fem-3* regulator) affected, while total lack of function induces, as expected for an essential splicing factor, multiple defects, leading to lethality.

In contrast to the apparent one-to-one relationship found for most DEAH genes in yeast and metazoans, the genes related to yeast YLR419w have largely diversified in metazoans (fig. 1). Moreover, a particular branch of this group, which includes *maleless* and its orthologs, is the only one among all DEAH helicases that has acquired new protein domains since the fungi/animals split. These results obviously make these DEAH metazoan genes good candidates for having diversified

their functions. The available evidence suggests that this group of genes are involved in metazoans in several, very different processes. Thus, Drosophila *spindle-E* (also known as *homeless*) is involved in RNA localization and probably translational regulation in the oocyte (Gillespie and Berg 1995; González-Reyes, Elliott, and St. Johnston 1997). The protein encoded by the Drosophila gene *maleless,* apart from its involvement in dosage compensation, may have a second function in resolving dsRNA structures prior to splicing, as suggested by recent data (Reenan, Hanrahan, and Ganetzky 2000; see also Kernan et al. 1991). Multiple functions have been described for the *maleless* ortholog in mammals, the gene encoding RNA helicase A (RHA; Lee and Hurwitz 1993; Zhang, Maacke, and Grosse 1995). Mouse *RHA* null mutants of both sexes die early in embryogenesis (Lee et al. 1998). It has been shown that RHA acts as a bridge for certain activators to associate to RNA polymerase II (Nakajima et al. 1997; Anderson et al. 1998; whether Drosophila MLE acts similarly in the context of the compensasome, to favor increased transcriptional levels, is unknown). Independently, human RHA has been involved in nuclear import/export, with the GGY C-terminal repeat acting as a nucleocytoplasmic shuttling signal (Tang et al. 1999; Westberg et al. 2000; for a review, see Michael 2000). The intriguing possibility that MLE is also involved in shuttling RNAs in and out of the nucleus cannot be dismissed. Recently, Eisen et al. (2000) characterized a Drosophila protein (DBP80) that interacted with MLE in a two-hybrid system. DBP80 turns to be the Drosophila ortholog of yeast DBP5 (our unpublished data), another RNA helicase specifically involved in RNA export

A

```
                                    *            20              *
Dm MOF       : VLQS TTENAAAPD-- YYVHYV LNRRLDGWVGRHR S : 37
Hs MOF       : VIQS VND-QE RE-- FYVHYV NRRLDEWVDKNR A : 36
Pw PWA33     : IIKT KTK--ENAE-- FYVHYV LNRRQNEWVDKSRV  : 35
Xl XNF7      : IVKT YNK-QA RE-- FYVHYV LNRRQNEWVDKSRL  : 36
Sc ESA1      : ILSINT R----KAPP--KFYVHYVN NKRLDEWITTDRI N : 34
Ce VC5.4     : VISRCRA--AN SI--KFYVHY IDCNRRLDEWVQSDRI N : 35
Hs TIP60     : ILSV DI---S RK--LFYVHYID NKRLDEWVTHERL D : 34
Dm EG0007.7  : IVSI EL---D RR-- FYVHYVD NKRLDEWVNEEDL Y : 34
Ce K03D10.3  : IENHEQG---AKEL--MYYVHYEELDRRNDEWITLDKI   : 34
Sp Alp13     : IVDTEM----K DV-TTYLIHYK KNSWDEWVEQDRI   : 34
Hs RBP1      : IKSTEID---D EV--LYLVHYY NVSYDEWVKADRI   : 34
Hs MSL-3     : IVVVIVGKDEK RKIP YLIHFN NRSWDRWAAEDHV   : 39
Dm MSL-3     : VLNVFERRNEH LRFY YKIHFQ RPSYDRCVRATVL   : 39
Dm MRG15     : VLKT PD---ATPV-- YYIHYA SKNWDEWVPENRV   : 34
Hs MRG15     : CVKVAIK---DKQV--KYFIHYS NKNWDEWVPESRV   : 34
Fr Sushi (RT): LLAV R----R RG-F YLVDWV GPEDQSWVPPSYL A : 34
At Legolas (RT): IMDRMTK-GTR KARDLLKVLWNCRGREEYTWETENKMK : 38
At Gimli (RT): ILKR LV-NRH RAATKVLVQWTNEDEAEATWEFLFDL   : 38
Dm PC        : IIQKEVK---K VV-- YRVKWK NQRYNTWEPEVNI   : 34
Mm M33       : ILSK LR---K KL-- YLVKWR SSKHNSWEPEENI   : 34
Hs PC3       : LLKR IR---K RM-- YLVKWK SQKYSTWEPEENI   : 34
Xl PC        : IEKK IR---K RV-- YLVKWR SSKYNTWEPEENI   : 34
Ce CEC1      : ILEH KK---K KS-- FYIKWL DHTHNSWEPKENI   : 34
Fo Skippy (RT): IRDT KI---N QR-- YLIKWKN PENENTWEPPKHL   : 34
Hs HP1       : VLDR VV---K KV-- YLLKWK SDEDNTWEPEENL D : 34
Dm HP1       : IIDR VR---K KV-- YYLKWK PETENTWEPENNL D : 34
```

B

0.1

(Snay-Hodge et al. 1998; Tseng et al. 1998). Of course, MLE also has a repetitious GGY C-terminus (Kuroda et al. 1991; see fig. 1). In summary, diversification of this group of proteins may have allowed a considerable sophistication of some aspects of RNA metabolism in metazoans with respect to yeasts.

A caveat is that we have been assuming in this discussion that yeasts have not lost any genes of this family. However, although it is not impossible that an *mle*-like gene (or genes) has been lost in yeasts, we think that such a possibility (which involves two independent events, duplication prior to the fungi/animals split and subsequent loss in the yeast lineage) is less parsimonious than the alternative (involving just gene duplication in the metazoan branch), especially considering the trend toward increased gene number in metazoans with respect to yeasts.

Evolution of MYST Acetyltransferases

Results shown in figures 3 and 5 suggest that the origin of the three groups of MYST acetyltransferases that we defined could be as follows. The presence of a chromodomain in proteins of two of these three groups suggests a first duplication giving rise to a *Sas3* gene plus a *Esa1/Sas2* gene. This was followed by acquisition of the chromodomain in the latter gene and a second duplication, producing the paralogous *Esa1* and *Sas2* genes. These two duplications occurred before the fungi/ animals split. Data shown in figures 3 and 4 suggest that duplications since the fungi/animals split have again occurred in two of the three groups, affecting the metazoan ortholog of *Sas3* and, at least in insects, the ortholog of *Sas2*. However, the time at which the *mof*/CG1894 duplication occurred cannot be precisely determined at present because the existence of more *mof*-related human genes cannot be ruled out.

Several previous authors have discussed the similarity of ESA1 to MOF, suggesting that they are orthologs (Allard et al. 1999; Galarneau et al. 2000). Data shown in figure 3, however, suggest that this is not the case, with the ortholog of *esa1* in Drosophila being the uncharacterized EG0007.7 gene. Purification of the yeast NuA4 and human TIP60 histone acetyltransferase complexes have shown that they contain several orthologous proteins (Allard et al. 1999; Galarneau et al. 2000; Ikura et al. 2000). Particularly, two members of the MYST family are present, ESA1 in the NuA4 complex and TIP60 in the TIP60 complex, which our analyses have shown to be orthologous. Our data thus predict that a related NuA4/TIP60 complex exists in Drosophila and that it contains EG0007.7. Unfortunately, data for complexes including the most likely orthologs of MOF in yeasts, SAS2, and human MOF protein are unavailable.

Figures 5 and 6 show that chromodomain and the PHD domains closely related to those found in MYST family members are present in other proteins. It is particularly interesting that the chromodomains in MSL-3 and its relatives are among the most similar to those found in MOF and its relatives, a result that can be explained by a common origin or by convergence due to functional interactions (e.g., the chromodomains of MSL-3 and MOF might interact during the assembly of the compensasome). It is also significant that the chromodomains most similar to those in MYST proteins so far available in the databases are found in two amphibian proteins (XNF7 and PWA33) that have orthologs in other vertebrate species that lack such domains (data not shown). This result suggests that the chromodomain has been recently acquired by those amphibian proteins and that it originated from a MYST family member. The close similarity of MYST PHD fingers to those in *Requiem* and its orthologs (fig. 6) is also striking. In this case, it is not possible to determine with the available data the origin of those domains.

Gene Duplication and the Origin of the Compensasome

In conclusion, our results show that, as shown for *msl-3* (Marín and Baker 2000), *mle* and *mof* have an ancient origin and that orthologs can be found in yeasts, nematodes, and humans. The possibility of a Drosophila-type compensasome complex existing in other organisms depends on the assumption that the lack of conservation at the sequence level for *msl-1* and *msl-2* is preventing the recognition of the orthologs of these two genes in other species. Even assuming this, we are skeptical about a Drosophila-like compensasome being formed in yeasts, considering the following: (1) The closest relatives of *msl-3, mof,* and *mle* in *S. cerevisiae* are quite different from the Drosophila proteins (Marín and Baker 2000; this study), including the lack of dsRNA-binding domains and GGY repeats in all yeast DEAH helicases and the lack of a chromodomain in the closest relative of *mof* in yeasts. (2) Duplicates of *msl-3* and *mle* exist that are more similar in sequence to the orthologous yeast proteins than the compensasome proteins themselves (Marín and Baker 2000; fig. 1). Those

---

←

FIG. 5.—Phylogenetic relationships among chromodomains of selected proteins. *A,* Multiple alignment of the most conserved region of the chromodomain (Aasland and Stewart 1995; Koonin, Zhou, and Lucchesi 1995). Apart from those found in MYST family proteins, we have included the canonical domains found in HP1 and POLYCOMB (PC) proteins, those in MSL-3 and some of its relatives (Marín and Baker 2000), some representatives of the chromodomains found in the integrases of a group of retrotransposons recently called *chromoviruses* (labeled "RT"; see Marín and Lloréns 2000), and the chromodomains that showed the highest similarity to those in MYST proteins in TBLASTN database searches (amphibian proteins PWA33 and XNF7). The two first letters are species name abbreviations. In addition to those species detailed in figure 1, this figure also includes data from the newt *Pleurodeles waltlii* (Pw), the frog *Xenopus laevis* (Xl), the yeast *Schizosaccharomyces pombe* (Sp), the fish *Fugu rubripes* (Fr), the plant *Arabidopsis thaliana* (At), the mouse *Mus musculus* (Mm), and the fungus *Fusarium oxysporum* (Fo). *B,* Phylogenetic tree showing the close relationship of MYST family protein chromodomains (highlighted in bold). Only bootstrap values greater than 600 are included.
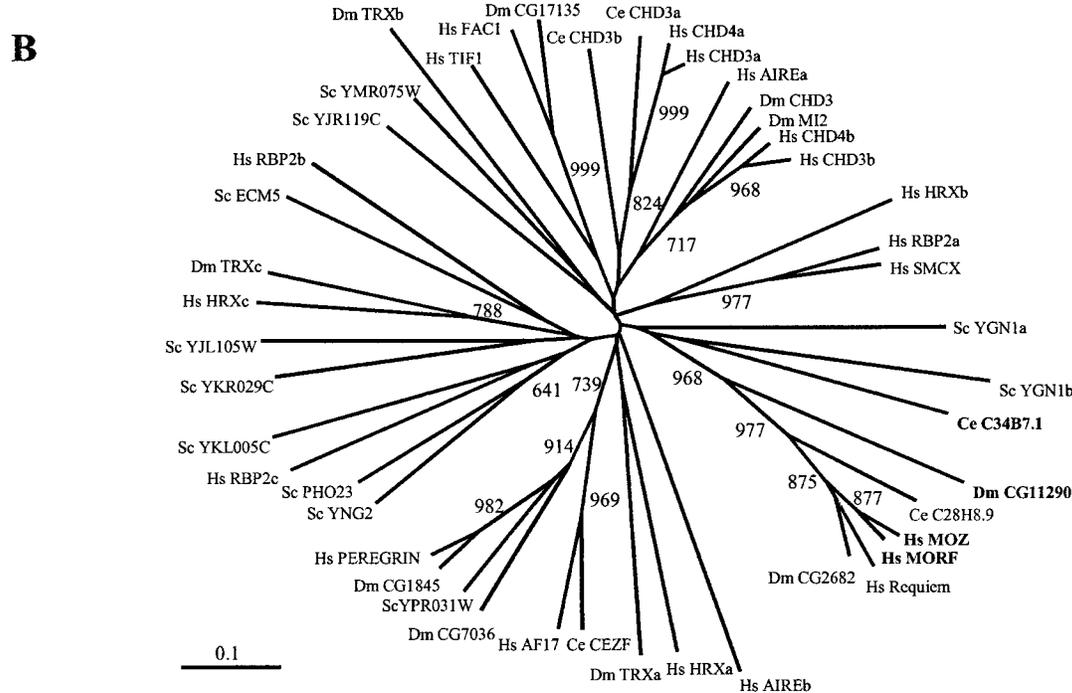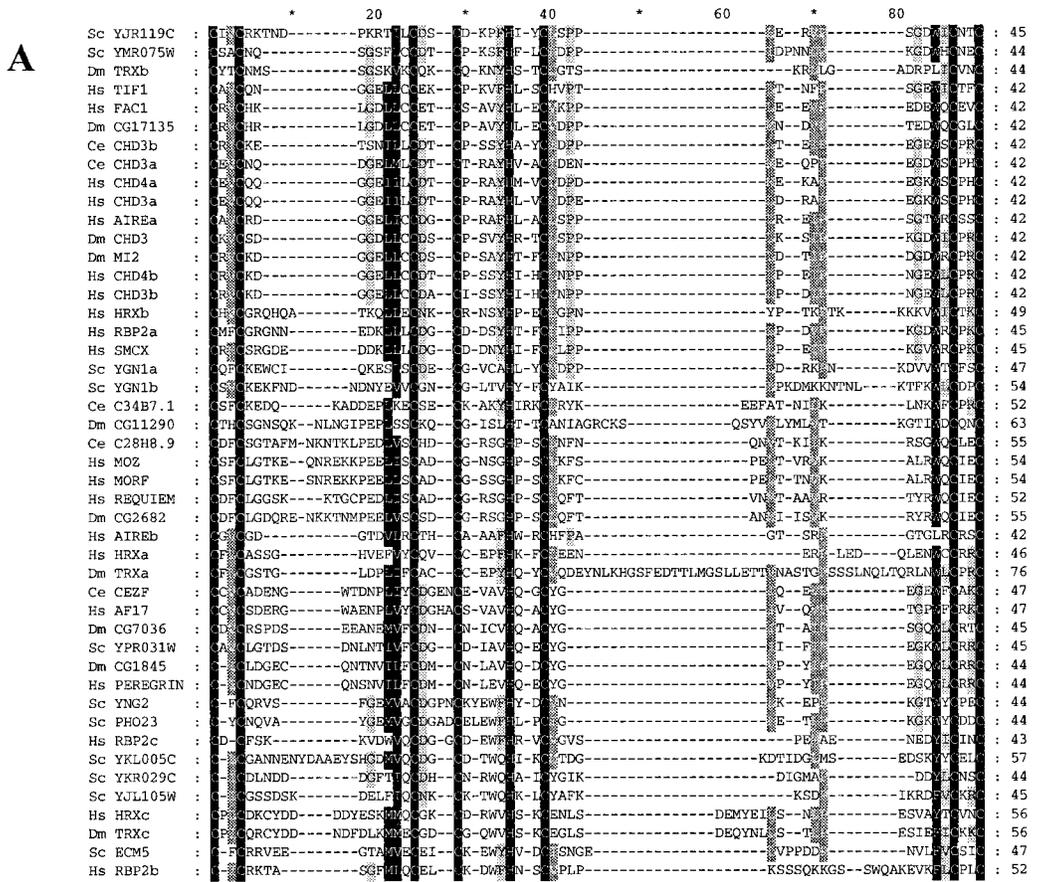
FIG. 6.—Phylogenetic relationships of selected PHD finger domains. *A,* Alignment based on results by Aasland, Gibson, and Stewart (1995) plus many other sequences found in database searches. The alignment includes data from only four species (Sc, Ce, Dm, Hs; see fig. 1). All PHD fingers with a close similarity to those found in MYST family proteins were included. Where more than one PHD finger was present in a protein, PHD fingers are distinguished by a small letter added at the end of the name. *B,* Phylogenetic tree, showing branches supported by bootstrap values greater than 600. As in the previous figure, MYST family proteins are shown in bold.

duplicates may have retained the original function found in yeast. It is also unlikely that the compensasome exists in *C. elegans,* considering that *msl-3* seems to have been lost in the nematode (Marín and Baker 2000). Therefore, the best candidates for having a Drosophila-like compensasome are mammalian species, for which very close relatives of *msl-3, mle,* and *mof* have been found. However, considering the interspecific variations observed for other chromatin-remodeling protein complexes (e.g., Wang et al. 1996; reviewed by Varga-Weisz and Becker 1998), it would not be surprising to find related complexes, containing some compensasome proteins combined with proteins not found in the Drosophila compensasome, in many species, including yeasts and nematodes.

We have shown that three compensasome genes have been involved in duplications since the fungi/animals split. Additionally, in the case of the MLE protein, acquisition of new protein domains has occurred. We propose that both gene duplication and cooption of new domains have been critical in the emergence of this protein complex. Recent data have highlighted the importance of gene duplication in the origin of complexity (Iwabe, Kuma, and Miyata 1996; Nadeau and Sankoff 1997). A particularly interesting model showed that the probability of conservation of duplicates was higher when duplication of several/many interacting genes occurred more or less simultaneously (Wagner 1994). If such multiple duplications occur, conservation of both copies is favored by one of the duplicates of each gene establishing new mutual interactions, while the other copy of each gene covers the original function. Such a model can be especially useful to understand the evolution of protein complexes. The prediction of such a model for a protein complex that arises de novo is that we should find coevolution of the proteins involved, with all of them arising by gene duplication at about the same time. The origin of the compensasome may fit such dynamics. Comparative analyses, particularly data for intermediate species, are necessary to test this hypothesis.

Acknowledgment

LITERATURE CITED

AASLAND, R. A., T. J. GIBSON, and A. F. STEWART. 1995. The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem. Sci. **20**:56–59.

AASLAND, R., and A. F. STEWART. 1995. The chromo shadow domain, a second chromo domain in heterochromatin-binding protein 1, HP1. Nucleic Acids Res. **23**:3168–3173.

AKHTAR, A., and P. B. BECKER. 2000. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in Drosophila. Mol. Cell **5**:367–375.

ALLARD, S., R. T. UTLEY, J. SAVARD, A. CLARKE, P. GRANT, C. J. BRANDL, L. PILLUS, J. L. WORKMAN, and J. CÔTÉ. 1999. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. EMBO J. **18**:5108–5119.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

ANDERSON, S. F., B. P. SCHLEGEL, T. NAKAJIMA, E. S. WOLPIN, and J. D. PARVIN. 1998. BRCA1 protein is linked to the RNA polymerase II holoenzyme complex via RNA helicase A. Nat. Genet. **19**:254–256.

BAKER, B. S., M. GORMAN, and I. MARÍN. 1994. Dosage compensation in *Drosophila.* Annu. Rev. Genet. **28**:491–521.

BASHAW, G. J., and B. S. BAKER. 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**:33245–3258.

BONE, J. R., J. LAVENDER, R. RICHMAN, M. J. PALMER, B. M. TURNER, and M. I. KURODA. 1994. Acetylated histone H4 on the male X chromosome is associated with dosage compensation in *Drosophila.* Genes Dev. **8**:96–104.

BORROW, J., V. P. STANTON JR., J. M. ANDRESEN et al. (15 coauthors). 1996. The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses a putative acetyltransferase to the CREB-binding protein. Nat. Genet. **14**:33–41.

BYCROFT, M., T. J. P. HUBBARD, M. PROCTOR, S. M. V. FREUND, and A. G. MURZIN. 1997. The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid–binding fold. Cell **88**:235–242.

CHAMPAGNE, N., N. R. BERTOS, N. PELLETIER, A. H. WANG, M. VEZMAR, Y. YANG, H. H. HENG, and X. J. YANG. 1999. Identification of a human histone acetyltransferase related to monocytic leukemia zinc finger protein. J. Biol. Chem. **274**:28528–28536.

CLINE, T. W., and B. J. MEYER. 1996. Vive la différence: males vs. females in flies and worms. Annu. Rev. Genet. **30**:637–702.

CONWAY MORRIS, S. 1998. Metazoan phylogenies: falling into place or falling to pieces? Curr. Opin. Gen. Dev. **8**:662–667.

COPPS, K., R. RICHMAN, L. M. LYMAN, K. A. CHANG, J. RAMPERSAD-AMMONS, and M. I. KURODA. 1998. Complex formation by the *Drosophila* MSL proteins: role of the MSL2 RING finger in protein complex assembly. EMBO J. **17**:5409–5417.

DE LA CRUZ, J., D. KRESSLER, and P. LINDER. 1999. Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. Trends Biochem. Sci. **24**:192–198.

EISEN, A., M. SATTAH, T. GAZITT, K. NEAL, P. SZAUTER, and J. LUCCHESI. 1998. A novel DEAD-box RNA helicase exhibits high sequence conservation from yeast to humans. Biochem. Biophys. Acta **1397**:131–136.

FRANKE, A., and B. S. BAKER. 1999. The *rox1* and *rox2* RNAs are essential components of the ''compensasome'', which mediates dosage compensation in *Drosophila.* Mol. Cell **4**:117–122.

———. 2000. Dosage compensation rox! Curr. Opin. Cell Biol. **12**:351–354.

GABIG, T. G., P. L. MANTEL, R. ROSLI, and C. D. CREAN. 1994. Requiem: a novel zinc finger gene essential for apoptosis in myeloid cells. J. Biol. Chem. **269**:29515–29519.

GALARNEAU, L., A. NOURANI, A. A. BOUDREAULT, Y. ZHANG, L. HÉLIOT, S. ALLARD, J. SAVARD, W. S. LANE, D. J. STILLMAN, and J. CÔTÉ. 2000. Multiple links between the NuA4 histone acetyltransferase complex and epigenetic control of transcription. Mol. Cell **5**:927–937.

GALLEGOS, M., J. AHRINGER, S. CRITTENDEN, and J. KIMBLE. 1998. Repression by the 3′ UTR of *fem-3,* a sex-determin-

ing gene, relies on a ubiquitous *mog*-dependent control in *Caenorhabditis elegans.* EMBO J. **17**:6337–6347.

GEE, S., S. W. KRAUSS, E. MILLER, K. AOYAGI, J. ARENAS, and J. G. CONBOY. 1997. Cloning of mDEAH9, a putative RNA helicase and mammalian homologue of *Saccharomyces cerevisiae* splicing factor Prp43. Proc. Natl. Acad. Sci. USA **94**:11803–11807.

GIBSON, T. J., and J. D. THOMPSON. 1994. Detection of Ds-RNA-binding domains in RNA helicase A and *Drosophila* Maleless: implications for monomeric RNA helicases. Nucleic Acids Res. **22**:2552–2556.

GILLESPIE, D. E., and C. A. BERG. 1995. Homeless is required for RNA localization in *Drosophila* oogenesis and encodes a new member of the DE-H family of RNA-dependent ATPases. Genes Dev. **9**:2495–2508.

GONZÁLEZ-REYES, A., H. ELLIOTT, and D. ST. JOHNSTON. 1997. Oocyte determination and the origin of polarity in *Drosophila*: the role of the *spindle* genes. Development **124**:4927–4937.

GRAHAM, P. L., and J. KIMBLE. 1993. The *mog-1* gene is required for the switch from spermatogenesis to oogenesis in *Caenorhabditis elegans.* Genetics **133**:919–931.

GRAHAM, P. L., T. SCHEDL, and J. KIMBLE. 1993. More *mog* genes that influence the switch from spermatogenesis to oogenesis in the hermaphrodite germ line of *Caenorhabditis elegans.* Dev. Genet. **14**:471–484.

GRIBSKOV, M. 1992. Translational initiation factor IF1 and factor eIF2a share an RNA binding motif with prokaryotic ribosomal protein S1 and polynucleotide phosphorylase. Gene **119**:107–111.

HILFIKER, A., D. HILFIKER-KLEINER, A. PANNUTI, and J. C. LUCCHESI. 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.

IIZUKA, M., and B. STILLMAN. 1999. Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J. Biol. Chem. **274**:23027–23034.

IKURA, T., V. V. OGRYZKO, M. GRIGORIEV, R. GROISMAN, J. WANG, M. HORIKOSHI, R. SCULLY, J. QIN, and Y. NAKATANI. 2000. Involvement of the TIP60 histone acetylase complex in DNA repair and apoptosis. Cell **102**:463–473.

IWABE N., K. KUMA, and T. MIYATA. 1996. Evolution of gene families and relationship with organismal evolution: rapid divergence of tissue-specific genes in the early evolution of chordates. Mol. Biol. Evol. **13**:483–493.

JANKOWSKY, E., and A. JANKOWSKY. 2000. The DExH/D protein family database. Nucleic Acids Res. **28**:333–334.

KELLEY, R. L., and M. I. KURODA. 2000. The role of chromosomal RNAs in marking the X for dosage compensation. Curr. Opin. Genet. Dev. **10**:555–561.

KELLEY, R. L., V. H. MELLER, P. R. GORDADZE, G. ROMAN, R. L. DAVIS, and M. I. KURODA. 1999. Epigenetic spreading of the *Drosophila* dosage compensation complex from *roX* RNA genes into flanking chromatin. Cell **98**:513–522.

KELLEY, R. L., I. SOLOVYEVA, L. M. LYMAN, R. RICHMAN, V. SOLOVYEV, and M. I. KURODA. 1995. Expression of *msl-2* causes assembly of dosage compensation regulators on the X chromosomes and female lethality in *Drosophila.* Cell **81**:867–877.

KERNAN, M. J., M. I. KURODA, R. KREBER, B. S. BAKER, and B. GANETZKY. 1991. *napts,* a mutation affecting sodium channel activity in *Drosophila,* is an allele of *mle,* a regulator of X chromosome transcription. Cell **66**:949–959.

KOONIN, E. V., S. ZHOU, and J. C. LUCCHESI. 1995. The chromo superfamily: new members, duplication of the chromo

domain and possible role in delivering transcription regulators to chromatin. Nucleic Acids Res. **23**:4229–4233.

KURODA, M. I., M. J. KERNAN, R. KREBER, B. GANETZKY, and B. S. BAKER. 1991. The maleless protein associates with the X chromosome to regulate dosage compensation in *Drosophila.* Cell **66**:935–947.

LEE, C.-G., K. A. CHANG, M. I. KURODA, and J. HURWITZ. 1997. The NTPase/helicase activities of *Drosophila* maleless, an essential factor in dosage compensation. EMBO J. **16**:2671–2681.

LEE, C.-G., V. DA COSTA SOARES, C. NEWBERGER, K. MANOVA, E. LACY, and J. HURWITZ. 1998. RNA helicase A is essential for normal gastrulation. Proc. Natl. Acad. Sci. USA **95**:13709–13713.

LEE, C.-G., and J. HURWITZ. 1993. Human RNA helicase A is homologous to the maleless protein of *Drosophila.* J. Biol. Chem. **268**:16822–16830.

LUCCHESI, J. C. 1998. Dosage compensation in flies and worms: the ups and downs of X-chromosome regulation. Curr. Opin. Genet. Dev. **8**:179–184.

MARÍN, I., and B. S. BAKER. 2000. Origin and evolution of the regulatory gene *male-specific lethal-3.* Mol. Biol. Evol. **17**:1240–1250.

MARÍN, I., A. FRANKE, G. J. BASHAW, and B. S. BAKER. 1996. The dosage compensation system of *Drosophila* is co-opted by newly evolved X chromosomes. Nature **383**:160–163.

MARÍN, I., and C. LLORÉNS. 2000. *Ty3/Gypsy* retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol. Biol. Evol. **17**:1040–1049.

MARÍN, I., M. SIEGAL, and B. S. BAKER. 2000. The evolution of dosage compensation mechanisms. Bioessays **22**:1106–1114.

MELLER, V. H. 2000. Dosage compensation: making 1X equal 2X. Trends Cell Biol. **10**:54–59.

MELLER, V. H., P. R. GORGADZE, Y. PARK, X. CHU, C. STUCKENHOLZ, R. L. KELLEY, and M. I. KURODA. 2000. Ordered assembly of *roX* RNAs into MSL complexes on the dosage compensated X chromosome in *Drosophila.* Curr. Biol. **10**:136–143.

MEYER, B. J. 2000. Sex in the worm. Counting and compensating X-chromosome dose. Trends Genet. **16**:247–253.

MICHAEL, W. M. 2000. Nucleocytoplasmic shuttling signals: two for the price of one. Trends Cell Biol. **10**:46–50.

NADEAU, J. H., and D. SANKOFF. 1997. Comparable rates of gene loss and functional divergence after genome duplications early in vertebrate evolution. Genetics **147**:1259–1266.

NAKAJIMA, T., C. UCHIDA, S. F. ANDERSON, C.-G. LEE, J. HURWITZ, J. D. PARVIN, and M. MONTMINY. 1997. RNA helicase A mediates association of CBP with RNA polymerase II. Cell **90**:1107–1112.

NEAL, K. C., A. PANNUTI, E. R. SMITH, and J. C. LUCCHESI. 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, K. B., and H. B. NICHOLAS, JR. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the authors (http://www.psc.edu/biomed/genedoc/).

ONO, Y., M. OHNO, and Y. SHIMURA. 1994. Identification of a putative RNA helicase (HRH1), a human homolog of yeast *Prp22.* Mol. Cell. Biol. **14**:7611–7620.

PAGE, R. D. M. 1996. TREEVIEW: an application to display phylogenetic tress on personal computers. CABIOS **12**:357–358.

PALMER, M. J., V. A. MERGNER, R. RICHMAN, J. E. MANNING, M. I. KURODA, and J. C. LUCCHESI. 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.

PUOTI, A., and J. KIMBLE. 1999. The *Caenorhabditis elegans* sex determination gene *mog-1* encodes a member of the DEAH-box protein family. Mol. Cell. Biol. **19**:2189–2197.

———. 2000. The hermaphrodite sperm/oocyte switch requires the *Caenorhabditis elegans* homologs of PRP2 and PRP22. Proc. Natl. Acad. Sci. USA **97**:3276–3281.

REENAN, R. A., C. J. HANRAHAN, and B. GANETZKY. 2000. The *mle^napts* RNA helicase mutation in *Drosophila* results in a splicing catastrophe of the *para* Na$^+$ channel transcript in a region of RNA editing. Neuron **25**:139–149.

REIFSNYDER, C., J. LOWELL, A. CLARKE, and L. PILLUS. 1996. Yeast SAS silencing genes and human genes associated with AML and HIV-1 Tat interactions are homologous with acetyltransferases. Nat. Genet. **14**:42–49.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SCOTT, M. J., L. L. PAN, S. B. CLELAND, A. L. KNOX, and J. HEINRICH. 2000. MSL1 plays a central role in assembly of the MSL complex, essential for dosage compensation in *Drosophila.* EMBO J. **19**:144–155.

SMITH, E. R., A. EISEN, W. GU, M. SATTAH, A. PANNUTI, J. ZHOU, R. G. COOK, J. C. LUCCHESI, and C. D. ALLIS. 1998. ESA1 is a histone aceteyltransferase that is essential for growth in yeast. Proc. Natl. Acad. Sci. USA **95**:3561–3565.

SMITH, E. R., A. PANUTTI, W. GU, A. STEURNAGEL, R. G. COOK, C. D. ALLIS, and J. C. LUCCHESI. 2000. The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol. Cell. Biol. **20**:312–318.

SNAY-HODGE, C. A., H. V. COLOT, A. L. GOLDSTEIN, and C. N. COLE. 1998. Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. EMBO J. **17**:2663–2676.

TANG, H., D. MCDONALD, T. MIDDLESWORTH, T. J. HOPE, and F. WONG-STAAL. 1999. The carboxyl terminus of RNA helicase A contains a bidirectional nuclear transport domain. Mol. Cell. Biol. **19**:3540–3550.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignments aided by quality analysis tools. Nucleic Acids Res. **24**:4876–4882.

TSENG, S. S., P. L. WEAVER, Y. LIU, M. HITOMI, A. M. TARTAKOFF, and T. H. CHANG. 1998. Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. EMBO J. **17**:2651–2662.

TURNER, B. M., A. J. BIRLEY, and J. LAVENDER. 1992. Histone H4 isoforms acetylated at specific lysine residues define individual chromosomes and chromatin domains in *Drosophila* polytene nuclei. Cell **69**:375–384.

VARGA-WEISZ, P. D., and P. B. BECKER. 1998. Chromatin-remodeling factors: machines that regulate? Curr. Opin. Cell Biol. **10**:346–353.

WAGNER, A. 1994. Evolution of gene networks by gene duplications: a mathematical model and its implications on genome organization. Proc. Natl. Acad. Sci. USA **91**:4387–4391.

WANG, W., Y. XUE, S. ZHOU, A. KUO, B. R. CAIRNS, and G. R. CRABTREE. 1996. Diversity and specialization of mammalian SWI/SNF complexes. Genes Dev. **10**:2117–2130.

WESTBERG, C., J.-P. YANG, H. TANG, T. R. REDDY, and F. WONG-STAAL. 2000. A novel shuttle protein binds to RNA helicase A and activates the retroviral constitutive transport element. J. Biol. Chem. **275**:21396–21401.

ZHANG, S., M. MAACKE, and F. GROSSE. 1995. Molecular cloning of the gene encoding nuclear DNA helicase II. J. Biol. Chem. **270**:16422–16427.

ZHOU, S., Y. YANG, M. J. SCOTT et al. (11 co-authors). 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.

ZHOU, Z., and R. REED. 1998. Human homologs of yeast prp16 and prp17 reveal conservation of the mechanism for catalytic step II of pre-mRNA splicing. EMBO J. **17**:2095–2106.