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.

Key words: dosage compensation, RNA helicase, histone acetyltransferase, chromodomain, PHD finger, comparative genomics.

Address for correspondence and reprints: Ignacio Marín, Departamento de Genética, Universidad de Valencia, Calle Doctor Moliner, 50, Burjassot 46100, Valencia, Spain. E-mail: ignacio.marin@uv.es.

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
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 acetylase 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.berkley.edu.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 mog-1, 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 (YLRS419w), 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
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, *D. melanogaster* CG11375/1405, and human genes encoding MOZ and MORF). Finally, *H. sapiens* HBO1 and *D. melanogaster* CG5229 contained an additional zinc finger domain of the C2HC 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
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-
Evolution of Compensosome Proteins

Fig. 2 (Continued)

The evolutionary relationships of the MYST family proteins were investigated. Consensus domains were identified in the sequences of various species, including S. cerevisiae (SAS2), D. melanogaster (CG1894), C. elegans (K03D10.3), and H. sapiens (MOF). These domains are found in other proteins as well, as shown in figures 5 and 6. The results indicate a monophyletic origin for the MYST family proteins.

The N-terminal chromodomains (fig. 4) were also analyzed. It was found that all chromodomains and PHD finger domains detected had a monophyletic origin. The results are shown in figures 5 and 6. As expected, proteins of the MYST family share high sequence similarity and structural characteristics, with three of them also having a chromodomain (fig. 4).

The N-terminal chromodomains of S. cerevisiae SAS2 and D. melanogaster CG1894, which also share 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
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 phylogenetic 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
gene function blocks the transition from sperm production to oocyte production in \textit{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 \textit{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 \textit{mog} genes indeed encode proteins involved solely in splicing and that the sex determination phenotypes of \textit{mog} mutations are indirectly caused by abnormal splicing of some regulator of the target sex determination gene \textit{fem-3} (as suggested by Puoti and Kimble 1999; see also Gallegos et al. 1998). In fact, worms lacking maternal \textit{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 \textit{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 \textit{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 \textit{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 is involved in metazoans in several, very different processes. Thus, Drosophila \textit{spindle-E} (also known as \textit{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 \textit{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 Gantetzky 2000; see also Kernan et al. 1991). Multiple functions have been described for the \textit{maleless} ortholog in mammals, the gene encoding RNA helicase A (RHA; Lee and Hurwitz 1993; Zhang, Maacke, and Grosse 1995). Mouse \textit{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.
detailed in figure 1, this figure also includes data from the newt database searches (amphibian proteins PWA33 and XNF7). The two first letters are species name abbreviations. In addition to those species (labeled "RT"; see Marõn and Lloreõns 2000), and the chromodomains that showed the highest similarity to those in MYST proteins in TBLASTN.

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 (Marõn 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 orthogonal yeast proteins than the compensosome proteins themselves (Marín and Baker 2000; fig. 1). Those

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 an 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 orthogonal 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 compensosome). 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 Compensosome

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 compensosome 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 compensosome 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 orthogonal yeast proteins than the compensosome proteins themselves (Marín and Baker 2000; fig. 1).
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**

We thank Bruce Baker for his comments on the manuscript.

**LITERATURE CITED**


Gallegos, M., J. Ahringer, S. Crittenden, and J. Kimble. 1998. Repression by the 3’T of *fem-3*, a sex-determin-


Evolution of Compensasome Proteins


DIETHARD TAUTZ, reviewing editor

Accepted November 9, 2000