Comparative Genomics of the RBR Family, Including the Parkinson's Disease–Related Gene *Parkin* and the Genes of the *Ariadne* Subfamily

Ignacio Marín and Alberto Ferrús†*

*Departamento de Genética, Universidad de Valencia, Burjassot, Valencia, Spain; and †Instituto Cajal, Consejo Superior de Investigaciones Científicas, Madrid

Genes of the RBR family are characterized by the RBR signature (two RING finger domains separated by an IBR/DRIL domain). The RBR family is widespread in eukaryotes, with numerous members in animals (mammals, *Drosophila*, *Caenorhabditis*) and plants (*Arabidopsis*). But yeasts, such as *Saccharomyces cerevisiae* or *Schizosaccharomyces pombe*, contain only two RBR genes. We determined the phylogenetic relationships and the most likely orthologs in different species of several family members for which functional data are available. These include: (1) *parkin*, whose mutations are involved in forms of familial Parkinson's disease; (2) the *ariadne* genes, recently characterized in *Drosophila* and mammals; (3) *XYbp* and *Dorfin*, two mammalian genes whose products interact with the centrosome; (4) *XAP3*, *RBCK1*, and *UIP28*, mammalian genes encoding Protein Kinase-C–binding proteins; and (5) *ARA54*, an androgen receptor coactivator. Because several of these genes are involved in ubiquitination, we used phylogenetic and structural analyses to explore the hypothesis that all RBR proteins might play a role in ubiquitination. We show that the involvement of RBR proteins in ubiquitination predates the animals-plants-fungi divergence. On the basis of the evidence provided by cases of gene fusion, we suggest that Ariadne proteins interact with cullin domain–containing proteins to form complexes with ubiquitin-ligase activity.

Introduction

Quantitative regulation of cell protein levels is a fine-tuned process whose biological relevance is now well established. In particular, the importance of protein ubiquitination is already understood in some detail. Conjugation of ubiquitin molecules with substrate proteins usually targets them for degradation via the 26S proteasome (Hershko 1997). Four types of proteins are known to mediate ubiquitination: ubiquitin-activating enzymes (referred to as E1 or Uba), ubiquitin-conjugating enzymes (E2 or Ubc), ubiquitin-protein ligases (E3 or Ubr), and the recently discovered polyubiquitin ligase (E4) (Koegl et al. 1999). These enzymes have been characterized in several organisms and most thoroughly in yeasts. In *Saccharomyces cerevisiae*, a single E1 protein exists, whereas the E2 proteins appear as a family of 11 sequence-related members (reviewed by Ciechanover 1998; Ciechanover, Orian, and Schwartz 2000). Multiple, functionally nonequivalent E2 enzymes are also found in other species.

Nonredundancy is shown by the fact that E2 mutations lead to lethality in Drosophila and mice (Harbers et al. 1996; Cenci et al. 1997) and to diverse phenotypes, including in several cases lethality, in yeasts (reviewed in Hochstrasser 1996). The study of ubiquitin ligases (E3s) reveals a further degree of complexity. At least six unrelated types of proteins or protein complexes exhibit E3 activity in diverse organisms. Nevertheless, a characteristic protein domain, the RING finger, appears in many proteins (e.g., CBL, BRCA-1, APC11, RBX1, etc.) able to act as E3 enzymes either by themselves or as part of multiprotein complexes (reviewed in Freemont 2000; Tyers and Jorgensen 2000). Considering that there are over 2,000 RING finger–containing proteins, it is not surprising that they have multiple roles, often unrelated to ubiquitination, and it is generally assumed that RING fingers mediate protein-protein interactions (reviewed in Borden and Freemont 1996). In ubiquitination, there is increasing evidence that the RING finger contributes, alone or in combination with other proteins (when the RING finger–containing protein is a part of an E3 complex), to interact with E2 enzymes. For example, in the ubiquitination complexes SCF and APC, RING fingers play a central role to determine substrate specificity (Tyers and Willems 1999). In all analyzed cases, an intact RING finger is essential for E3 function (see reviews by Jackson et al. 2000; Joazeiro and Weissman 2000).

Mutations in several human E3s are involved in pathogenesis. Particularly, the gene *parkin* is mutated in some cases of autosomal recessive familial Parkinson's disease (Kitada et al. 1998; Polymeropoulos 2000). Parkin is an E3 enzyme (Shimura et al. 2000; Zhang et al. 2000) that contains a characteristic cysteine-rich region formed by three consecutive protein domains, namely two C_3HC_4 RING fingers (Klug and Schwabe 1995; Saurin et al. 1996) and a peculiar C_6HC domain separating them. This C_6HC domain has been called IBR (''in-between-RINGS'') finger (Morett and Bork 1999) or DRIL (''double RING finger–linked'') domain (Van der Reijden et al. 1999). In this study we refer to the group of genes that encode proteins with a RING-IBR/DRIL-RING signature as the RBR (RING-Between rings-RING) family.

RBR family genes are widespread in eukaryotes, including humans (Morett and Bork 1999; Van der Reijden et al. 1999; Aguilera et al. 2000). Some members of this family have been studied in detail, and the available data suggest that not only Parkin but also other RBR proteins are involved in ubiquitin metabolism. Thus, the products of the *Drosophila ariadne-1* and *ariadne-2* genes have been shown to interact with ubiquitin-conjugating enzymes (Aguilera et al. 2000), an in-

Key words: evolution, gene fusion, ubiquitination, ubiquitin ligase.

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.

Mol. Biol. Evol. 19(12):2039–2050. 2002

teraction also demonstrated for mammalian Ariadnelike proteins (Martinez-Noel et al. 1999; Moynihan et al. 1999). Similarly, three RBR mouse genes called *UIP4*, *UIP28*, and *UIP117* were discovered in a yeast two-hybrid screen for proteins interacting with the mouse ubiquitin-conjugating enzyme UbcM4 (Martínez-Noel et al. 1999). *UIP4* is a putative ortholog of a human gene of unknown function (*KIAA0161*) (Martinez-Noel et al. 1999). *UIP28* shows very significant similarity with the rat gene *RBCK1* (Tokunaga et al. 1998) and the human gene *Xap3* (Cong et al. 1997). *UIP117*, also known as *XYbp*, has been more extensively studied. It is expressed in high levels in the testis, where it concentrates in pachytene spermatocytes (Párraga and del Mazo 2000). At the cellular level, the XYbp protein localizes in centrosomes and the XY body (Párraga and del Mazo 2000). The product of a close relative of *XYbp* in humans, called *Dorfin*, is also a constituent of the centrosome and interacts with an E2 enzyme (Niwa et al. 2001). *UIP117/XYbp* shows substantial similarity with *mgp*, a gonadal-specific gene of the mosquito *Aedes aegypti* (Zhao et al. 2000).

In this study we performed a comprehensive analysis of the RBR family to determine the origin and meaning of the RBR signature and the pattern of diversification of RBR genes. We also establish the origin of the involvement of RBR proteins in ubiquitination, suggest functional roles for some of these proteins, and determine the orthologs in several eukaryotic species of human RBR genes of known or potential medical interest.

Materials and Methods

Criteria Used to Screen for RBR Family Members

Van der Reijden et al. (1999) defined the C_6HC DRIL domain and named TRIAD proteins as those containing a RING-DRIL-RING signature. Independently, Morett and Bork (1999) defined the IBR domain that is essentially equivalent to the DRIL domain. Definition of the IBR domain, however, was less strict than that of the DRIL domain; consequently, some proteins (e.g., Parkin, RBCK1) were defined either as RING-IBR-RING proteins (Morett and Bork 1999) or as just having a signature related to, but different from that found in the RING-DRIL-RING proteins (i.e., they cannot be considered TRIAD proteins; Van der Reijden et al. 1999). To avoid this type of discrepancies we followed an open search strategy. We hypothesized that it would be very unlikely that many different, independently arisen, signatures could mimic what is in summary a $C_3HC_{10}HC_4HC_4$ series of conserved residues. Thus, as a first approximation we assumed that all cysteine-histidine patterns resembling a $C_3HC_{10}HC_4HC_4$ signature must be evolutionarily related to each other, indicating that the genes that carry them belong to the RBR family.

As a starting point, we used the ''nonredundant'' and ''month'' protein databases found online in the National Center for Biotechnology Information web pages, (NCBI; http://www.ncbi.nlm.nih.gov/) to perform multiple PSI BLAST searches (Altschul et al. 1997) using as seeds all the proteins described by Morett and Bork (1999) and Van der Reijden et al. (1999). We tentatively accepted as RBR family members all proteins detected in those searches that showed signatures with conserved cysteine and histidine residues that could correspond to a RING-IBR/DRIL-RING arrangement. In the second step we used the sequences of the proteins found in those preliminary searches to perform TBLASTN (Altschul et al. 1997) analyses to detect all putative RBR proteins not present in the protein databases. The ''nonredundant,'' ''month,'' ''dBEST,'' and ''htgs'' nucleotide databases at NCBI were screened. We then pooled all the known or putative proteins found and made preliminary multiple alignments (see *Multiple-Sequence Alignments*). In a third step we used the output of those preliminary alignments for the following purposes: (1) To eliminate duplicates. (2) To eliminate a few sequences that had, on close inspection, only a low similarity with the rest. In particular, we discarded some proteins that were picked up in our searches because of their similarity with RBR proteins but lacked many of the conserved histidine or cysteine residues that define the RING finger or the IBR/DRIL domain. (3) Whenever possible we eliminated from the sequences obtained from genomic sources those regions that obviously corresponded to introns. We thus obtained a database of proteins with a conserved pattern of cysteine and histidine residues that could correspond to a RING-IBR/DRIL-RING signature, albeit without making a strong selection to eliminate those that had one or few cysteine or histidine residues missing. These analyses were completed in December 2001. After this general screening, we validated that all those sequences had at least one functional RING finger domain by comparing the pattern of cysteine and histidine residues in the putative RING finger regions with previous descriptions of the RING finger domain (reviewed in Borden and Freemont 1996). Particularly, we established that the patterns found did not correspond to LIM domains or PHD fingers, domains that are relatively similar to the RING finger (Borden and Freemont 1996), using the InterPro (http://www.ebi.ac.uk/interpro/) entries for these alternative domains (LIM domain: IPR001781; PHD finger: IPR001965). For RING finger comparisons detailed below, a database of 276 RING finger–containing proteins was selected among those found in the SMART database (http://smart.embl-heidelberg.de; Smart accession number SM0184).

Multiple-Sequence Alignments

Both for preliminary screening of the NCBI databases (see above) and after the database of RBR proteins was completed, we used the default parameters of the ClustalX 1.81 program (Thompson et al. 1997) to obtain alignments of the RBR signature. The alignment of the final database was refined manually, editing the sequences with the GeneDoc program (Version 2.5; Nicholas KB and Nicholas HB 1997). GeneDoc was also used to highlight the similarities among sequences. Similar procedures were used to establish alignments of both

the N- and C-terminal RING fingers in RBR proteins with the RING finger sequences obtained from SMART.

Phylogenetic Analyses

Phylogenetic trees were obtained from the protein multiple alignments, using three different approaches. First, we used the Neighbor-joining (NJ) method (Saitou and Nei 1987), as implemented in ClustalX. To determine the reliability of the topology of the NJ tree, we performed 1,000 bootstrap replicates using the routine available in ClustalX also. Second, we used maximum parsimony (MP), as implemented in MEGA 2.1 (Kumar et al. 2001). Parameters used were: (1) all sites included, (2) heuristic search based on close-neighbor interchange with search level 3, and (3) initial trees obtained by random addition with 10 replicates. Following Nei and Kumar (2000; p. 131) we present data for the bootstrap consensus tree, also obtained using MEGA 2.1. Finally, PROTML 2.2, included in the package Molphy 2.2 (Adachi and Hasegawa 1992; version compiled for PC computers by Professor Russell L. Malmberg, University of Georgia, USA. Available at http://dogwood.botany.uga.edu/malmberg/software.html), was used to obtain maximum-likelihood–based trees. The random addition (−q) option with the default JTT-f amino acid replacement model was used to obtain the 50 best trees. Those trees were evaluated using the approximate bootstrapping method known as ''resampling of estimated log likelihood'' (RELL; Kishino, Miyata, and Hasegawa 1990) to obtain the values presented below. Trees were drawn using TreeView 1.6.1 (Page 1996).

Structural Analyses

We performed TBLASTN and PSI BLAST analyses using as query the sequences outside the RBR signature of all RBR family proteins available at the NCBI protein databases to determine whether other conserved domains existed in those proteins. These analyses are useful not only to confirm the relationships among the members of the RBR family but also to determine whether RBR genes have fused or shuffled domains with genes outside the RBR family.

Results

Ancient Eukaryotic Origin of the RBR Family

Using the procedures detailed above, we obtained an exhaustive representation of all proteins with a strict pattern of conserved residues C_3HC_4–C_6HC–C_3HC_4 while including proteins that lack one or few of these residues as well. We excluded the possibility of any of those cysteine-rich sequences corresponding to the LIM or PHD domains, quite similar to RING fingers (Borden and Freemont 1996), following the accepted definitions of those domains (see InterPro database). This strategy thus allows the detection of both typical and diverged (i.e., those with some unusual residues) family members, thereby establishing which residues in the main RBR signature may be occasionally dispensable for protein function.

According to those criteria we found a large number of potential RBR genes in eukaryotic species. Many sequences of animals, plants, and fungi, and a few sequences from protozoans, such as *Plasmodium falciparum* (GenBank accession number AL034558), *Leishmania major* (GenBank accession number AL352992), *Entamoeba histolytica* (GenBank accession numbers AZ681233, AZ690910), or *Trypanosoma brucei* (GenBank accession number AC009259), and from the slime mold *Dictyostelium discoideum* (GenBank accession number C91572), were detected. No candidates were obtained from eubacteria or archaea species. Therefore, we concluded that the RBR signature appeared early in eukaryotic evolution. We aimed then to determine the evolutionary history of the RBR family in crown group eukaryotes for which the available genomic information is extensive. In particular, we took advantage of the complete or substantially complete genomic sequence data for several animal, plant, and yeast species, which allowed a comprehensive study of gene emergence, conservation, and change over the last billion years (Knoll 1992; Doolittle et al. 1996). To this end, we aligned all the RBR family genes found in two yeast species (*S. cerevisiae* and *Schizosaccharomyces pombe*), a plant (*Arabidopsis thaliana*), and four animals (*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, and *Caenorhabditis elegans*). A total of 74 sequences were aligned. The alignment is available at the EBI alignment database (ftp:/ftp.ebi.ac.uk/pub/databases/embl/align) with accession number ALIGN_000336.

Differential Features of the Two RING Fingers in RBR Proteins

Even assuming that changes from histidine to cysteine (or vice versa) in the critical positions of the RING and IBR/DRIL domains may be allowed, we found that 33 of the 74 proteins lacked one or several of those characteristic residues. Some residues are particularly often altered. Thus, the characteristic histidine in the C-terminal RING finger domain is substituted by diverse amino acids in 28 proteins. In summary, 25 of the 33 proteins have uncommon residues only in the C-terminal RING finger, whereas just three proteins have changes only in the N-terminal RING finger and two show a change in the IBR/DRIL domain. More than two changes are found only in one protein (*Arabidopsis* T20M3.15, three changes). This protein, however, is very similar to other RBR proteins, so it is clearly a member of the RBR family (see *Phylogenetic Analysis of the RBR Family*).

Another interesting result that emerges from the multiple sequence alignment is the substantial difference in spacing between conserved residues and in length variability between the two RING finger domains. The first RING finger has a general formula $C–X_2–C–X_{(10–24)}–C–X_{(1–6)}–H–X_2–C–X_2–C–X_{(14–25)}–C–X_{(2–9)}–C$. It is thus rather long and substantially variable in length. The second RING finger has, with a single exception, a much shorter and strict formula: $C–X_2–C–X_{(9–11)}–C–X–H–X_2–C–X_{(1–4)}–C–X_4–C–X_2–C$. The exception is that a

very long region (32 amino acids) is found in the *Drosophila* protein (CG11321) between the second and the third cysteines of its second RING finger. This exception may be the result of this region corresponding to a small intron in the *Drosophila* gene. It is quite unlikely, however, that introns could explain the greater variability or longer extension of the first RING finger compared with that of the second RING finger because similar patterns are observed in several RBR proteins.

These differences between the two RING fingers are likely to be functionally significant. We established an alignment among both RING finger domains in RBR proteins and those in the 276 proteins obtained from the SMART database (see *Materials and Methods*) and checked for conserved residues (defined as such when at least 40% of the proteins contained identical or biochemically similar amino acids). Apart from the characteristic cysteine and histidine residues, a total of 13 other conserved positions were found in the alignment, in agreement with those found by Borden and Freemont (1996) and with the consensus sequences established in the SMART database. Inspection of both RING fingers in RBR proteins, however, showed that although the characteristic amino acids were found in at least 11 of those 13 conserved positions in many N-terminal RING fingers of RBR proteins, only three of those 13 were consistently found in their C-terminal RING fingers (data not shown). Therefore, we can conclude that the C-terminal RING finger often lacks in one of the most conserved cysteine and histidine residues and has an amino acid composition that is quite distinct from that found in other RING fingers, including the RBR N-terminal one.

Phylogenetic Analysis of the RBR Family

The three independent methods of phylogenetic reconstruction generated very similar trees; these are summarized in figure 1. Although MP provided, in general, lower bootstrap values for the topology shown in figure 1, the high congruence among the three procedures offers good support for phylogenetic inference.

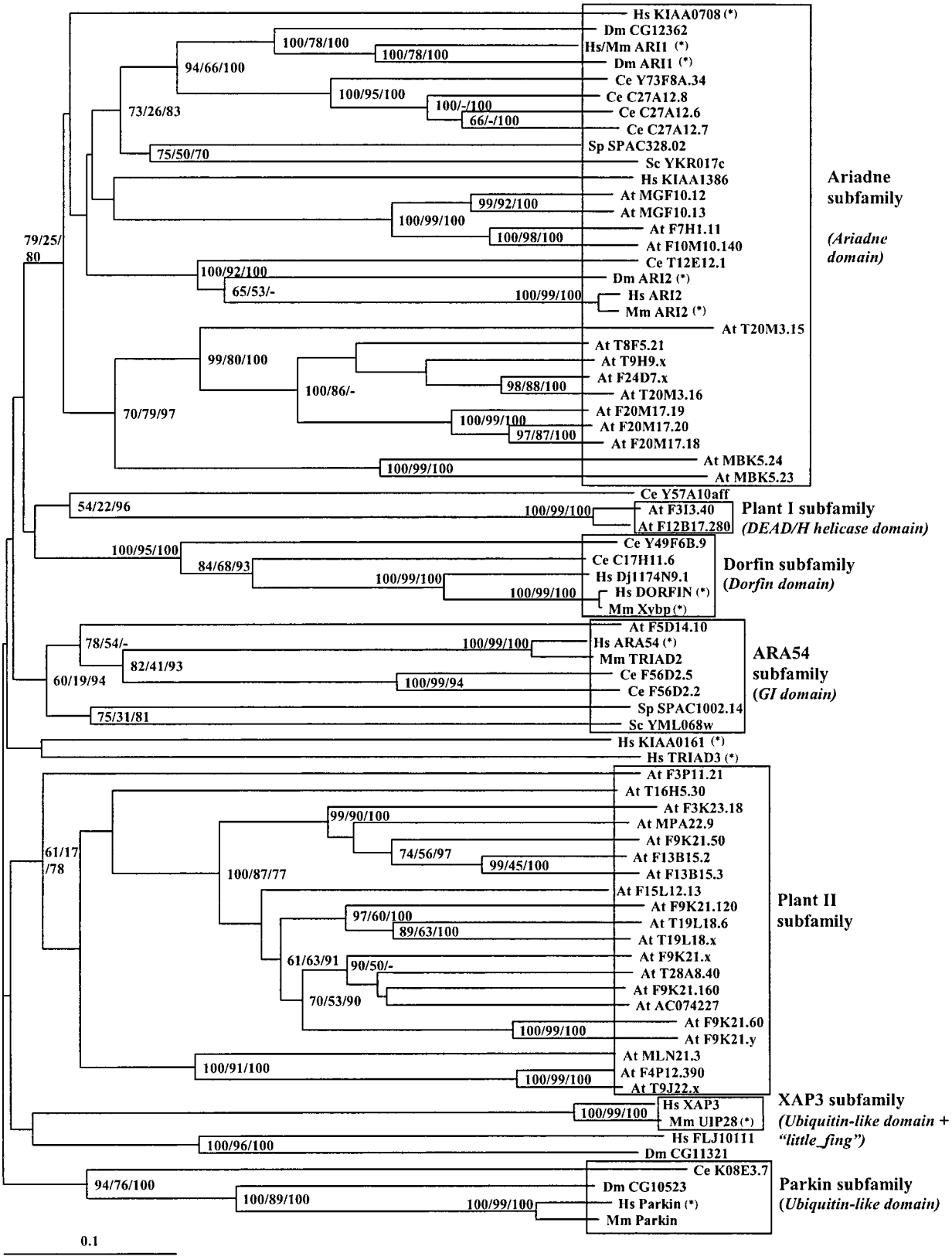
The first conclusion obtained from the phylogenetic data is that genes from animals, plants, or fungi form several groups. Plant genes are found in five distant positions in the tree. Similarly, the two RBR genes found in yeast species are far apart from each other in the tree. Animal genes form even more disjunct groups. At least six well-supported groups with putative orthologs in both protostomes and deuterostomes are found. These results strongly suggest that the RBR family diversified very early into different subfamilies, in many cases before the plants-fungi-animals splits.

A group, supported by the three methods of tree reconstruction, that contains genes from plants, animals, and fungi includes the *ariadne* genes (fig. 1). These results suggest that *ariadne* already formed a subfamily, independent from the rest of the RBR family genes, before all the considered species diversified. Another interesting finding is the existence of clear candidates for orthologs of mammalian *parkin* in invertebrate species but not in plants or fungi (Parkin subfamily; fig. 1). Also noteworthy is the finding of several functionally uncharacterized or poorly understood human genes (*Triad3*, *KIAA0161*, *FLJ10111*, *XAP3*) that appear scattered in this tree, their relative positions being uncertain (see fig. 1). The other known human RBR gene, *Dorfin*, appears in an independent group (Dorfin subfamily). *Dorfin* putative orthologs are found not only in mouse (*UIP117/XYbp*) but also in *C. elegans* (*Y49F6B.9* and *C17H11.6*). If the *mgp* gene of the mosquito *Aedes* is included in the tree, it becomes a part of the monophyletic subfamily formed by these mammalian and nematode genes (data not shown). It is thus surprising that no *Dorfin* ortholog has been detected in *Drosophila*. An interesting finding involving the Dorfin subfamily emerged when we evaluated human RBR genes as potential candidates to be associated with familial Parkinson's disease. In a recent report, Valente et al. (2001) detected a locus (PARK6) involved in autosomal recessive juvenile parkinsonism, the same pathology related to *parkin* mutations. The PARK6 locus mapped at chromosome 1, region 1p35–1p36. We have found that a human paralog of *Dorfin*, provisionally called *Dj1174N9.1* (fig. 1), maps at 1p34–1p35.3; thus, it may be a candidate for the PARK6 locus.

The human gene *ARA54* was characterized as an androgen receptor coactivator (Kang et al. 1999). The data (fig. 1) show that this gene belongs to the RBR family, being part of the only group apart from the Ariadne subfamily, for which we have support for orthologs existing in animals (*ARA54* in humans; *Triad2* in mouse; *F56D2.2* and *F56D2.5* in *C. elegans*), plants (*F5D14.10* in *A. thaliana*), and fungi (*SPAC1002.14* in *S. pombe* and *YML068w* in *S. cerevisiae*). We have named this group as ARA54 subfamily.

→

FIG. 1.—Neighbor-joining (NJ) phylogenetic tree of RBR family proteins. Maximum parsimony (MP) and maximum-likelihood (ML) trees were very similar; so all data are presented together. Numbers indicate bootstrap support (NJ/MP/ML) for each branch in percentages (for details see *Materials and Methods*). Values are shown only when at least two of these methods provided values over 50%. When a particular value for a branch is not shown, being substituted by a dash, it means that the branch was not supported by the corresponding method of reconstruction. The two first letters of the sequences names indicate the species (Hs: *Homo sapiens*; Mm: *Mus musculus*; Dm: *Drosophila melanogaster*; Ce: *Caenorhabditis elegans*; Sc: *Saccharomyces cerevisiae*; Sp: *Schizosaccharomyces pombe*; At: *Arabidopsis thaliana*). *Arabidopsis thaliana* proteins that have not been annotated in the *Arabidopsis* genome project are referred to by the name of the clone plus a letter (x, y). Human and mouse *Ariadne-1* have identical RBR signatures and are thus included only once under the name Hs/Mm_ARI1. Subfamilies and their characteristic domains (if present; in italics) are indicated. Asterisks (*) refer to those proteins that interact with the human E2 enzymes UbcH7, UbcH8, or their mouse and Drosophila ortholog proteins. ARA54 also interacts with UbcH6 and UbcH9. Data for interactions are from Martinez-Noel et al. (1999), Moynihan et al. (1999), Aguilera et al. (2000), Shimura et al. (2000), Zhang et al. (2000), Ito et al. (2001), and Niwa et al. (2001).


Hs KIAA0708 (*)
Dm CG12362
Hs/Mm ARI1 (*)
Dm ARI1 (*)
Ce Y73F8A.34
Ce C27A12.8
Ce C27A12.6
Ce C27A12.7
Sp SPAC328.02
Sc YKR017c
Hs KIAA1386
At MGF10.12
At MGF10.13
At F7H1.11
At F10M10.140
Ce T12E12.1
Dm ARI2 (*)
Hs ARI2
Mm ARI2 (*)
At T20M3.15
At T8F5.21
At T9H9.x
At F24D7.x
At T20M3.16
At F20M17.19
At F20M17.20
At F20M17.18
At MBK5.24
At MBK5.23
Ariadne subfamily
(Ariadne domain)
Ce Y57A10aff
At F3I3.40
At F12B17.280
Plant I subfamily
(DEAD/H helicase domain)
Ce Y49F6B.9
Ce C17H11.6
Hs Dj1174N9.1
Hs DORFIN (*)
Mm Xybp (*)
Dorfin subfamily
(Dorfin domain)
At F5D14.10
Hs ARA54 (*)
Mm TRIAD2
Ce F56D2.5
Ce F56D2.2
Sp SPAC1002.14
Sc YML068w
ARA54 subfamily
(GI domain)
Hs KIAA0161 (*)
Hs TRIAD3 (*)
At F3P11.21
At T16H5.30
At F3K23.18
At MPA22.9
At F9K21.50
At F13B15.2
At F13B15.3
At F15L12.13
At F9K21.120
At T19L18.6
At T19L18.x
At F9K21.x
At T28A8.40
At F9K21.160
At AC074227
At F9K21.60
At F9K21.y
At MLN21.3
At F4P12.390
At T9J22.x
Plant II subfamily
Hs XAP3
Mm UIP28 (*)
XAP3 subfamily
(Ubiquitin-like domain + "little_fing")
Hs FLJ10111
Dm CG11321
Ce K08E3.7
Dm CG10523
Hs Parkin (*)
Mm Parkin
Parkin subfamily
(Ubiquitin-like domain)
79/25/80
94/66/100
100/78/100
100/78/100
73/26/83
100/95/100
100/-/100
66/-/100
75/50/70
99/92/100
100/99/100
100/98/100
100/92/100
65/53/-
100/99/100
99/80/100
100/86/-
98/88/100
70/79/97
100/99/100
97/87/100
100/99/100
54/22/96
100/99/100
100/95/100
84/68/93
100/99/100
100/99/100
78/54/-
100/99/100
82/41/93
60/19/94
100/99/94
75/31/81
99/90/100
74/56/97
99/45/100
61/17/78
100/87/77
97/60/100
89/63/100
61/63/91
90/50/-
70/53/90
100/99/100
100/91/100
100/99/100
100/99/100
100/96/100
94/76/100
100/89/100
100/99/100
0.1

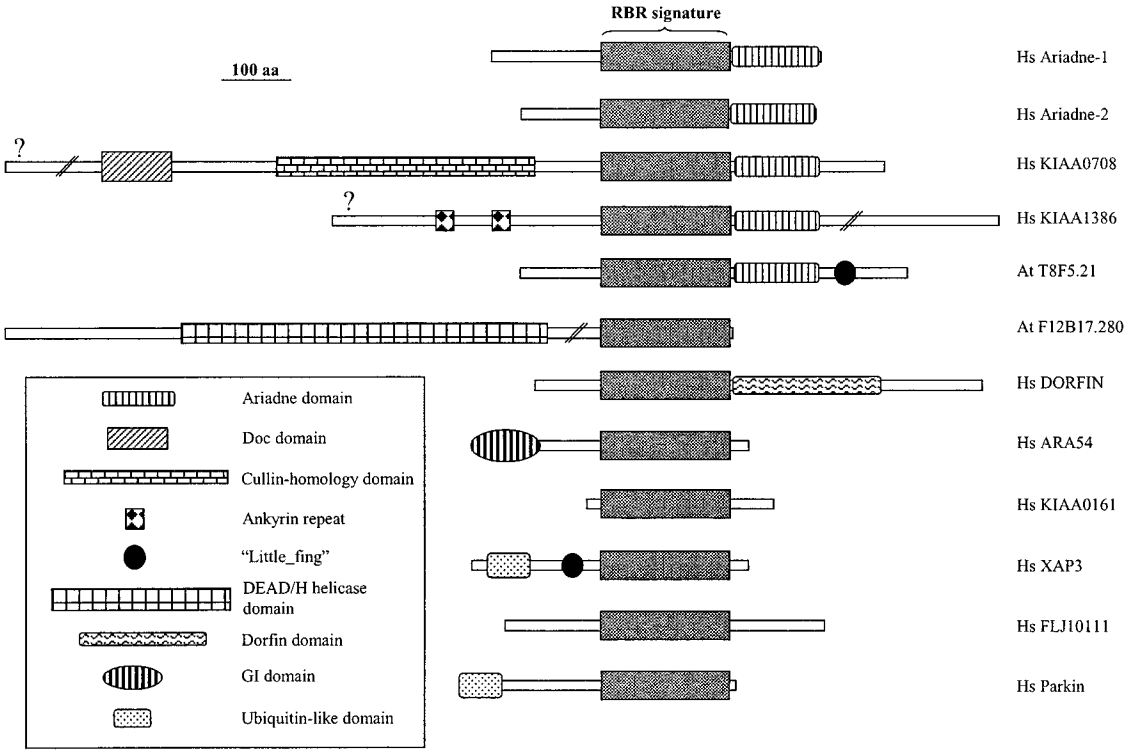


FIG. 2.—Structures of significant RBR family proteins. Question marks refer to uncertainties in the determination of N-terminal ends. Some proteins (KIAA0708, KIAA1386, F12B17.280) are too long to be represented in this scale and a noninformative portion of each of them has been eliminated from this schematic drawing. See text for details on the nature of these protein domains.

Another significant finding is the large diversification of some RBR subfamilies in plants. Two groups are particularly numerous in *Arabidopsis*. In contrast to the four human Ariadne subfamily genes, there are 14 in *A. thaliana*. A second likely monophyletic group (Plant II subfamily, see fig. 1) has no less than 20 members. Finding several very similar members of these subgroups positioned closely in the *Arabidopsis* genome suggests that, in part, these large numbers may be explained by recent tandem duplications (e.g., there are six very similar genes in a single *A. thaliana* clone, F9K21. See also pairs and trios in fig. 1).

Figure 1 also includes data of positive interactions of RBR proteins with Ubcs. The fact that the two very similar human proteins UbcH7 and UbcH8 (or their orthologs in Drosophila and mouse) interact with RBRs located all along the phylogenetic tree suggests that interaction was established before the diversification of many of the RBR proteins. This result suggests that most RBR proteins could interact with the same E2s and, therefore, be a part of an identical ubiquitination system in all the species in which they are present.

Structural Analyses Confirm the Phylogenetic Tree and Suggest that a Few Mammalian and Plant RBR Genes Emerged by Gene Fusion or Domain Shuffling

The main results of our structural analyses of the 74 sequences (see *Materials and Methods*) are summarized in figures 1 and 2. They are fully compatible and strongly support the phylogenetic tree because only closely related proteins, forming groups already detected by the three phylogenetic reconstruction procedures, have distinctive domains apart from the RBR signature (fig. 1). Examples of the different structures found are summarized in figure 2. All those genes that had, according to the phylogenetic analysis, a high similarity with the characterized *ariadne* gene shared a second region, C-terminal with respect to the RBR signature. This region is 150 to 165 amino acids long and corresponds to residues 329–492 in *Drosophila* Ariadne-1 protein. It has no similarity with previously described domains, so we have called this characteristic region as ''Ariadne domain'' (fig. 2). Phylogenetic and structural analyses thus support the existence of the Ariadne subfamily. Although most Ariadne subfamily genes have very similar sizes and can be aligned satisfactorily along their whole lengths, two human genes of this subfamily (*KIAA1386* and *KIAA0708*) are, respectively, about 600 and >1,000 acids longer than the rest. A precise analysis showed that they both have peculiar structures. *KIAA1386* contains two ankyrin repeats, not detected in any other RBR protein (fig. 2). The structure of *KIAA0708* is even more interesting because it suggests that this gene emerged by the fusion of two preexisting genes (figs. 2 and 3). *KIAA0708* was originally described (Ishikawa et al. 1998) as a 5′ truncated cDNA (Accession number BAA31683). A N-terminally truncated, 1,753 amino acids–long protein is encoded by that cDNA. We found that the C-terminal end of the truncated KIAA0708 protein (amino acids 1,233–1,721 in the BAA31683 clone)

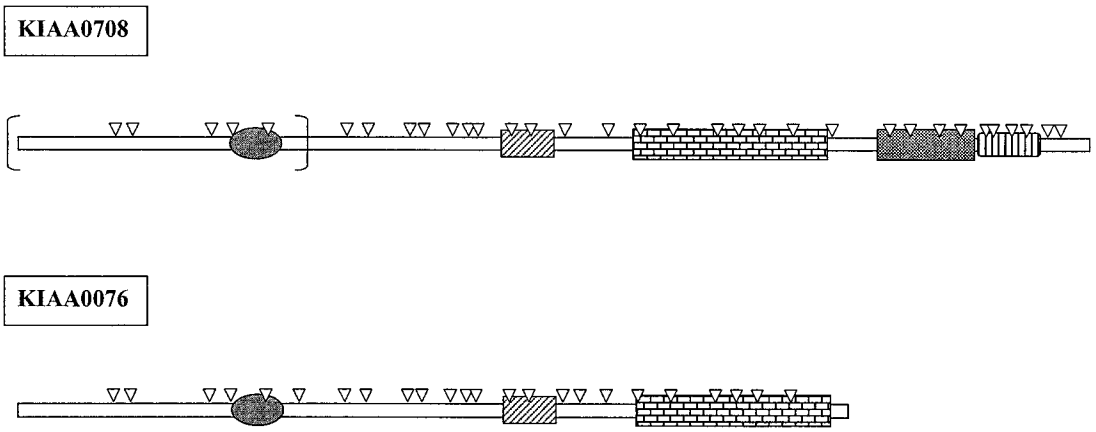


FIG. 3.—Structures of the human *KIAA0708* and *KIAA0076* genes. As explained in the text, a part of the structure of the KIAA0708 protein was determined using the information provided by a long cDNA (Accession number BAA31683) and confirmed by analyses of other cDNAs, whereas the most N-terminal end of the protein (in parenthesis) was deduced by comparison of genomic DNA with the KIAA0076 protein. Domains as in figure 2, with the exception of the most N-terminal domain that corresponds to a specific region of similarity to HERC2 proteins (see text). The small triangles show the position of introns in both human genes.

shows strong similarity (E values up to 10^{-42}) with other Ariadne subfamily proteins (figs. 1 and 2). But amino acids 154–1,190 show a high similarity (63% identity; E value = 0) with a protein encoded by a gene that apparently lacks an RBR domain. This gene, called *KIAA0076* (Nomura et al. 1994), encodes another very long protein (1,698 amino acids). The similarity with the truncated KIAA0708 spans from amino acids 589 to 1,649 in KIAA0076, reaching the C-terminus. When the region common to KIAA0076 and the truncated KIAA0708 proteins is analyzed, two well-known domains called Doc (Grossberger et al. 1999) and cullin-like (Yu et al. 1998) are found. These domains are characteristic of some proteins involved in ubiquitination, as a part of E3 complexes (reviewed in Grossberger et al. 1999; Jackson et al. 2000; Tyers and Jorgensen 2000). In addition to these two domains, KIAA0076 also contains a third, uncharacterized domain (amino acids 337–433), found also in the HERC2 proteins from humans, mouse, and Drosophila (Ji et al. 2000). Whether KIAA0708 contains this third domain or not cannot be established using the sequence in the BAA31683 clone. But TBLASTN searches against human genomic sequences detected homology to KIAA0076 in a region upstream of the beginning of the sequence found in the BAA31683 clone. We thus used that homology to reconstruct the likely complete sequence of the human *KIAA0708* gene, finding that it should also contain the HERC2-related domain (fig. 3). These results strongly suggest that *KIAA0708* emerged by the fusion of an Ariadne subfamily gene and a *KIAA0076*-like gene. That such fusion is not an artifact of the BAA31683 clone is demonstrated by the fact that we found other human cDNAs (Accession numbers BG034881 and BE896423) that contain, as BAA31683, the end of the cullin-like domain together with the beginning of the Ariadne-related sequence. Moreover, the exon-intron structure of *KIAA0708* and *KIAA0076* genes strongly suggests that they must have arisen by a relatively recent duplication because introns of both genes are found in almost identical positions (fig. 3). Finally, the fact that *KIAA0708* and *KIAA0076* are located very close to each other in the human chromosome 6, about 260-kb apart, suggests a tandem duplication of *KIAA0076* associated with a fusion with an Ariadne subfamily gene as the most likely origin of *KIAA0708*.

Mammalian *Xap3, UIP28,* and *RBCK1* genes may also have emerged by gene fusion. A recent report (Lim et al. 2001) points out the close similarity of a region present in these genes with the rat *Sharpin* gene, that encodes a Shank-interacting protein (Ehlers 1999). But the authors called this region ''RH domain,'' when in fact the similarity corresponds to a Ubiquitinlike domain plus a characteristic C4 zinc finger (named ''little_fing'' in the ''LOAD'' database of protein motifs developed by Aravind, Sreekumar, and Koonin and accessible through NCBI searches). The very high similarity of *Xap3/UIP28/RBCK1* and *Sharpin* (E value: 10^{-18}) suggests that those mammalian genes emerged by fusion of a *Sharpin*like gene, or a part of it, with an RBR family gene. Interestingly, both the little_fing C4 zinc finger domain and the Ubiquitinlike domain are present in other RBR proteins. It is unclear whether these domains have a common ancestry or have been co-opted twice independently by RBR proteins. Thus, a single *Arabidopsis ariadne* gene, *T8F5.21,* has a little_fing, although this time it was located C-terminally with respect to the RBR and Ariadne domains (fig. 2). Similarly, *Parkin* and its orthologs also contain a Ubiquitin-like domain, although it is very divergent from the one in *XAP3* and its relatives.

Another interesting feature present in some RBR proteins is the protein-protein interaction domain known as GI (Kubota, Sakaki, and Ito 2000). This domain is present in all genes in the *ARA54* subfamily (see fig. 1), with the possible exception of the *S. cerevisiae* gene *YML1068w* in which only a low degree of similarity has been found. This domain was previously found in proteins, such as the one encoded by the mammalian imprinted gene *Impact* and its orthologs in other animal

and yeast species (Hagiwara et al. 1997; Okamura et al. 2000), the protein kinase GCN2 (also from yeast or animals; Kubota, Sakaki, and Ito 2000) and the mouse AO7 protein. The latter is a RING finger protein known to interact with ubiquitin-conjugating enzymes (Lorick et al. 1999; Kubota, Sakaki, and Ito 2000).

Finally, *Arabidopsis* genes *F3I3.40* and *F12B17.280* encode very large proteins, 2,322 and 1,751 amino acids, respectively. Their structures also suggest the occurrence of gene fusion or domain shuffling (fig. 2). In this case, a 600-amino acid–long domain, located N-terminally with respect to the RBR domain, has strong similarity with DEAD/H helicases. Their position in the tree, independent of other plant genes, and their peculiar structures suggest that these genes may form a separated group, that we have named Plant I subfamily.

Discussion

Origin and Evolution of the RBR Family

Our findings strongly suggest that the RBR signature is an ancient eukaryotic feature and that several of the subfamilies of the RBR family originated before the plants versus animals-fungi split, around 1 billion years ago. There are five types of RBR proteins in plants, and considering the topology of the tree it is likely that genes belonging to several, perhaps all, of those types were already present when plants separated from the animals-fungi lineage. The two RBR genes characterized in yeasts are also so far apart in the tree and so similar to several plant and animal genes that it is logical to conclude that they emerged before the plants-animals-fungi splits. Furthermore, the phylogenetic analysis suggests that genes of at least six RBR subfamilies were already present before the protostome-deuterostome separation. The topology of the tree is compatible with five of those subfamilies being present before the plants-animals-fungi divergence. Taken together, these results support the hypothesis that genes of several, perhaps no less than five, RBR subfamilies were present in the last common ancestor of plants, animals, and fungi, with yeasts losing several of the genes later.

It is not unusual to find in a family genes that have acquired new protein domains. These domains may allow for the diversification of their biochemical functions. Our structural analyses showed that this phenomenon occurred several times during an RBR family evolution (figs. 1 and 2). For example, a recent acquisition of ankyrin repeats by an *ariadne* gene generated the peculiar *KIAA1386* gene. Similarly, *parkin* subfamily genes have acquired an ubiquitinlike domain in their N-terminus. Less frequent, however, is the fusion of two genes to become a single hybrid gene. *KIAA0708* may be one of those rare examples. Our analyses suggest that this gene may have emerged by the duplication (perhaps a tandem duplication) of a gene such as *KIAA0076*, followed by the fusion of this duplicate with a closely located *ariadne*like gene. Other cases of gene fusion may have occurred (see *Results*).

Evolution and Function of RBR Subfamilies

We discuss subsequently the different subfamilies defined according to our phylogenetic and structural analyses.

Parkin subfamily

Human *parkin* has orthologs in other vertebrates and in protostomes (insects, nematodes) but apparently not in yeasts or plants. This result suggests that the origin of *parkin* is relatively recent, although loss of *parkin* orthologs in plants and yeasts cannot be excluded. There is no available information on the functions of the protostome orthologs of *parkin.* Although it is tempting to suggest that they may also function as E3 enzymes, the lack of several key residues in the first RING finger of the *C. elegans* ortholog, *K08E3.7,* serves—if indeed true, and not caused by sequencing artifacts—as a warning for a possible functional divergence from their mammalian counterpart. The finding of several uncharacterized genes that have substantial similarity in their RBR domains with *parkin* (as the human genes *FLJ10111, KIAA0161, TRIAD3,* or *XAP3*) suggests related functions. In fact, it is known that the murine orthologs of *KIAA0161* and *XAP3* interact with E2 proteins (Martinez-Noel et al. 1999). Moreover, XAP3 has, as does Parkin, a Ubiquitinlike domain. They are therefore good candidates for being involved in pathologies akin to Parkinson's disease. But none of these genes localized in regions where Parkinson's disease genes have been mapped thus far (Polymeropoulos 2000; Valente et al. 2001). Therefore, although we can hypothesize that they are E3s, their precise cellular role may be different from the one performed by Parkin. In any case, the high similarity of XAP3 with a Shank-interacting protein, Sharpin, suggests a role in the nervous system, in particular in postsynaptic membranes (Lim et al. 2001). Considering the increasing evidence that links the ubiquitin pathway to neural connectivity, at least in *Drosophila* (Muralidhar and Thomas 1993; Oh et al. 1994; DiAntonio et al. 2001), it would be interesting to determine whether Parkin plays a role in a similar context. In fact, Parkin has been shown to mediate ubiquitin-dependent degradation of the synaptic vesicle–associated protein CDCrel-1 (Zhang et al. 2000).

Ariadne subfamily

It is a clear example of an RBR subfamily that predates the origin of plants, animals, and fungi. It may be significant that (excluding the exceptional human genes *KIAA0708* and *KIAA1386* that we will discuss below) all animal *ariadnes* belong to just two different groups, corresponding to *Drosophila ariadne-1* and *ariadne-2* genes (Aguilera et al. 2000) and their respective orthologs. Yeast *ariadnes* are significantly more similar to animal *ariadne-1* genes than to the *ariadne-2* group. It is thus possible that two *ariadnes* were already present before the animals and fungi diverged and *ariadne-2* was subsequently lost in fungi. Plant *ariadnes* also form two distinct groups. But the topology of the tree does

not allow to determine whether these groups correspond, respectively, to *ariadne-1* and *ariadne-2* genes that later became duplicated in plants. The two groups of plant genes are similarly related to both animal *ariadnes.*

Only the *KIAA0708* and *KIAA1386* human genes have an abnormal phylogenetic position within the Ariadne subfamily. They appear to be more related to plant *ariadnes* than to any other animal gene. But the finding that these are also the only two Ariadne subfamily genes with peculiar structures (see *Results* and figs. 2 and 3) suggests that their RBR domain sequences may have suffered selective pressure to adapt to their new protein contexts, perhaps causing them to evolve faster than normal and thus making their phylogenetic position within the Ariadne subfamily ambiguous.

Considering the conservation of Ariadne proteins in very different species, we would expect all Ariadne subfamily genes to perform related functions. A reasonable working hypothesis is that they all encode, as does *parkin,* ubiquitin ligases (E3s). For several members of the subfamily, interactions with Ubc (E2) enzymes have been demonstrated (see fig. 1). The gene fusion involving *KIAA0708,* however, casts doubts about Ariadnes being independent E3s. That fusion (fig. 3) put together an Ariadne subfamily gene with a second one, *KIAA0076,* that is similar to members of E3 complexes (i.e., it has Doc and cullin-homology domains). Cullin domain–containing proteins generally associate with several other proteins, including one containing a RING finger, to function as an E3 complex. This occurs in the SCF, VBC, and APC complexes (reviewed in Jackson et al. 2000). Because a gene fusion tends to be favored when two proteins physically interact (the basis for the ''Rossetta stone'' strategy; Marcotte et al. 1999), we suggest that Ariadnes are a part of the E3 complexes that contain a cullin domain protein, instead of functioning as independent E3s.

A final consideration with respect to Ariadne subfamily genes concerns the finding of a duplicate of *ariadne-1* in *D. melanogaster* (the gene provisionally called CG12362, see fig. 1). Null mutations in *ariadne-1* yield a few adult escapers, whereas those in *ariadne-2* are fully penetrant lethals (Aguilera et al. 2000). The finding of a recent *ariadne-1* duplicate suggests that this difference may be attributed to the partially redundant functions among the *ariadne-1* paralogs. Potential redundancy should be considered in all future analyses of RBR genes because of the extensive number of close duplicates detected.

Dorfin subfamily

The monophyletic group that contains mammalian *Dorfin* and *UIP117/XYbp* also has protostome representatives. The available data suggest that these genes might encode proteins also involved in ubiquitin metabolism (Martinez-Noel et al. 1999), mostly, but not exclusively, in reproductive tissues (Párraga and del Mazo 2000; Zhao et al. 2000; Niwa et al. 2001). Their localization on the centrosomes and XY body has a precedent in SCF ubiquitin-ligase complexes (Freed et al. 1999). The finding that a human paralog of *Dorfin* (*Dj1174N9.1*) maps where a locus (PARK6) involved in autosomal recessive early-onset Parkinson's disease has been located (Valente et al. 2001) suggests the convenience to analyze this gene for a potential involvement in that disease.

XAP3 subfamily

Because of their characteristic structures we provisionally define this subfamily as including a single mammalian gene (*XAP3* in humans, *UIP28* in mouse, *RBCK1* in rat). But the existence of distant relatives in protostomes or even in plants (Plant II subfamily) cannot be excluded (fig. 1). UIP28 protein interacts with E2s (Martinez-Noel et al. 1999), again suggesting a role in ubiquitination. How ubiquitination may be related to Protein kinase C interaction (Cong et al. 1997; Tokunaga et al. 1998) or to the ability of the murine protein to activate transcription by interacting with hepatitis B virus X protein (Cong et al. 1997) is intriguing. The similarity with *Sharpin* (see *Results*) suggests a role for *XAP3* and its orthologs in ubiquitinating proteins at the postsynaptic membrane.

ARA54 subfamily

This is a very likely monophyletic group that includes *ARA54* along with RBR genes in plants, fungi, and other animals. The presence of a characteristic protein domain (GI domain) in most, perhaps all, *ARA54* group proteins suggests specific functional features. Members of this subfamily may play a role in transcriptional activation, as documented for *ARA54* (Kang et al. 1999). The protein encoded by the putative ortholog of *ARA54* in *S. cerevisiae* (*YML068w*) interacts in a two-hybrid system with SNF11, a member of the SWI-SNF complex involved in chromatin remodeling (Uetz et al. 2000).

Additional genes in plants

All yeast and animal RBR genes are included in one of the previous five classes. But there are no less than 20 genes in *Arabidopsis* that form an additional, likely monophyletic, group (Plant II subfamily). Their closest animal relatives are *parkin* and some of its related genes (especially human *Xap3* and *FLJ1011,* mouse *UIP28,* and Drosophila *CG11321;* see fig. 1), precisely the only animal genes for which no potential orthologs are observed in plants. But the degree of divergence is such that it cannot be reliably determined whether these animal and plant genes are orthologous. Finally, we have found two exceptionally large plant genes (*F3I3.40* and *F12B17.280;* Plant I subfamily) whose phylogenetic position is also uncertain and, in addition, they have a peculiar helicase domain (fig. 2). The presence of this domain suggests again characteristic functions for these two proteins.

In conclusion, there is strong evidence that RBR proteins are involved in ubiquitin metabolism, potentially acting as E3 enzymes or members of an E3 com-

plex, since before the plants-animals-fungi divergence. This conclusion rests on the fact that proteins all along the phylogenetic tree interact with E2s enzymes. The diverse protein motifs that appear in those proteins may then have contributed to the acquisition of specialized functions in different tissues or cellular contexts.

Functional Considerations

Genes missing one or a few of the amino acids that characterize the RING or IBR/DRIL domains are interspersed with those that have the complete signature. Thus, it is obvious that a restrictive definition of the RBR family would have caused many members to be eliminated from our analyses. This result not only vindicates the strategy followed in this study but also has functional significance. Loss of critical residues in many RBR proteins means that the complete signature may often be unnecessary for protein function. In particular, whereas only very few proteins have alterations in the critical residues that define their N-terminal RING fingers, more than 30% of them have suffered changes in the critical residues of their C-terminal RING finger domain. Thus, the presence of a canonical second RING finger is not required in many RBR proteins for their role in ubiquitination. We have also found that both RING fingers have quite a different length and the C-terminal one has a quite uncommon amino acid composition. These results strongly suggest that both RING fingers are not functionally equivalent.

Indeed, many RING finger–containing proteins in E3 complexes have only one RING finger domain. An attractive hypothesis is that RBR proteins function as independent E3s only when both RING fingers are present, whereas their recruitment into E3 complexes relaxes the functional constraints imposed on one (most often the C-terminal) RING finger. Interestingly, mutations in both RING fingers of the independent E3 Parkin cause loss or difficulty in binding to E2s (Imai, Soda, and Takahashi 2000; Shimura et al. 2000; Zhang et al. 2000). In contrast, Ariadne proteins, that we propose are a part of E3 complexes, interact with the E2 enzyme only through their N-terminal RING finger domain (Moynihan et al. 1999; Aguilera et al. 2000; Ardley et al. 2001). All Ariadne proteins have intact N-terminal RING fingers, but loss of critical residues in the C-terminal domain is observed in eight of the 29 members examined. This result suggests that the second RING finger could be dispensable in some Ariadne proteins. Those that conserve an intact C-terminal RING finger may still require it for a function independent from Ubc interaction. This is the case of *Drosophila* Ariadne-1, where a mutation affecting the conserved C-terminal RING finger has a lethal phenotype, albeit that the motif does not play a role in Ubc interaction (Aguilera et al. 2000).

Data for most positive RBR-E2 interactions involve two very similar human proteins, UbcH7 and UbcH8, and their mouse and Drosophila orthologs. Only ARA54 has been shown to interact also with the less-related proteins UbcH6 and UbcH9 in a yeast two-hybrid assay (Ito et al. 2001). For the other RBR proteins, a few other Ubcs were tested with negative results. These data, considered at face value, suggest that all ubiquitination-involved RBRs would interact with a particular type of Ubc protein and that such an interaction was established preceding the diversification of the RBR family. But it must be considered that these experiments are still very incomplete because only a few of the many E2 enzymes that exist in any eukaryotic species have been tested so far.

Acknowledgments

This work was supported by grants PM 99-0099 from the DGICYT (Spanish Ministry of Culture) and 8.5/0043/1998 from the Community of Madrid awarded to A.F and by grant 01/080-00 from Fundació La Caixa. We appreciate the comments from lab members.

LITERATURE CITED

ADACHI, J., and M. HASEGAWA. 1992. Computer science monographs, Vol. 27. MOLPHY: programs for molecular phylogeny I-PROTML: maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo.

AGUILERA, M., M. OLIVEROS, M. MARTÍNEZ-PADRÓN, J. A. BARBAS, and A. FERRÚS. 2000. *Ariadne-1:* a vital *Drosophila* gene is required in development and defines a new conserved family of RING-finger proteins. Genetics **155**: 1231–1244.

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.

ARDLEY, H. C., N. G. TAN, S. A. ROSE, A. F. MARKHAM, and P. A. ROBINSON. 2001. Features of the Parkin/Ariadne ubiquitin ligase, HHARI, which regulate its interaction with the ubiquitin-conjugating enzyme, UbcH7. J. Biol. Chem. **276**: 19640–19647.

BORDEN, K. L. B., and P. S. FREEMONT. 1996. The RING finger domain: a recent example of a sequence-structure family. Curr. Opin. Struct. Biol. **6**:395–401.

CENCI, G., R. B. RAWSON, G. BELLONI, D. H. CASTRILLON, M. TUDOR, R. PETRUCCI, L. L. GOLDBERG, S. A. WASSERMAN, and M. GATTI. 1997. UbcD1, a *Drosophila* ubiquitin-conjugating enzyme required for proper telomere behavior. Genes Dev. **11**:863–875.

CIECHANOVER, A. 1998. The ubiquitin-proteasome pathway: on protein death and cell life. EMBO J. **17**:7151–7160.

CIECHANOVER, A., A. ORIAN, and A. L. SCHWARTZ. 2000. Ubiquitin-mediated proteolysis: biological regulation via destruction. Bioessays **22**:442–451.

CONG, Y. S., Y. L. YAO, W. M. YANG, N. KUZHANDAIVELU, and E. SETO. 1997. The hepatitis B virus X-associated protein, XAP3, is a protein kinase C-binding protein. J. Biol. Chem. **272**:16482–16489.

DIANTONIO, A., A. P. HAGHIGHI, S. L. PORTMAN, J. D. LEE, A. M. AMARANTO, and C. S. GOODMAN. 2001. Ubiquitination-dependent mechanisms regulate synaptic growth and function. Nature **412**:449–452.

DOOLITTLE, R. F., D. F. FENG, S. TSANG, G. CHO, and E. LITTLE. 1996. Determining divergence times of the major kingdoms of living organisms with a protein clock. Science **271**: 470–477.

EHLERS, M. D. 1999. Synapse structure: glutamate receptors connected by the shanks. Curr. Biol. **9**:R848–R850.

FREED, E., K. R. LACEY, P. HUIE, S. A. LYAPINA, R. J. DESHAIES, T. STEARNS, and P. K. JACKSON. 1999. Components of an SCF ubiquitin ligase localize to the centrosome and regulate the centrosome duplication cycle. Genes Dev. **13**: 2242–2257.

FREEMONT, P. S. 2000. RING for destruction? Curr. Biol. **10**: R84–R87.

GROSSBERGER, R., C. GIEFFERS, W. ZACHARIAE, A. V. PODTELEJNIKOV, A. SCHLEIFFER, K. NASMYTH, M. MANN, and J. M. PETERS. 1999. Characterization of the DOC1/APC10 subunit of the yeast and the human anaphase-promoting complex. J. Biol. Chem. **274**:14500–14507.

HAGIWARA, Y., M. HIRAI, K. NISHIYAMA, I. KANAZAWA, T. UEDA, Y. SAKAKI, and T. ITO. 1997. Screening for imprinted genes by allelic message display: identification of a paternally expressed gene impact on mouse chromosome 18. Proc. Natl. Acad. Sci. USA **94**:9249–9254.

HARBERS, K., U. MILLER, A. GRAMS, E. LI, R. JAENISCH, and T. FRANZ. 1996. Provirus integration into a gene encoding a ubiquitin-conjugating enzyme results in a placental defect and embryonic lethality. Proc. Natl. Acad. Sci. USA **93**: 12412–12417.

HERSHKO, A. 1997. Roles of ubiquitin-mediated proteolysis in cell cycle control. Curr. Opin. Cell Biol. **9**:788–799.

HOCHSTRASSER, M. 1996. Ubiquitin-dependent protein degradation. Annu. Rev. Genet. **30**:405–439.

IMAI, Y., M. SODA, and R. TAKAHASHI. 2000. Parkin suppresses unfolded protein stress-induced cell death through its E3 ubiquitin-protein ligase activity. J. Biol. Chem. **275**:35661–35664.

ISHIKAWA, K., T. NAGASE, M. SUYAMA, N. MIYAJIMA, A. TANAKA, H. KOTANI, N. NOMURA, and O. OHARA. 1998. Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro. DNA Res. **5**:169–176.

ITO, K., S. ADACHI, R. IWAKAMI, H. YASUDA, Y. MUTO, N. SEKI, and Y. OKANO. 2001. N-terminally extended human ubiquitin-conjugating enzymes (E2s) mediate the ubiquitination of RING-finger proteins, ARA54 and RNF8. Eur. J. Biochem. **268**:2725–2732.

JACKSON, P. K., A. G. ELDRIDGE, E. FREED, L. FURSTENTHAL, J. Y. HSU, B. K. KAISER, and J. D. R. REIMANN. 2000. The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases. Trends Cell Biol. **10**:429–439.

JI, Y., N. A. REBERT, J. M. JOSLIN, M. J. HIGGINS, R. A. SCHULTZ, and R. D. NICHOLLS. 2000. Structure of the highly conserved HERC2 gene and of multiple partially duplicated paralogs in human. Genome Res. **10**:319–329.

JOAZEIRO, C. A. P., and A. M. WEISSMAN. 2000. RING finger proteins: mediators of ubiquitin ligase activity. Cell **102**: 549–552.

KANG, H. Y., S. YEH, N. FUJIMOTO, and C. CHANG. 1999. Cloning and characterization of human prostate coactivator ARA54, a novel protein that associates with the androgen receptor. J. Biol. Chem. **274**:8570–8576.

KISHINO, H., T. MIYATA, and M. HASEGAWA. 1990. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J. Mol. Evol. **31**:151–160.

KITADA, T., S. ASAKAWA, N. HATTORI, H. MATSUMINE, Y. YAMAMURA, S. MINOSHIMA, M. YOKOCHI, Y. MIZUNO, and N. SHIMIZU. 1998. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Nature **392**:605–608.

KLUG, A., and J. W. SCHWABE. 1995. Protein motifs 5. Zinc fingers. FASEB J. **9**:597–604.

KNOLL, A. H. 1992. The early evolution of eukaryotes: a geological perspective. Science **256**:622–627.

KOEGL, M., T. HOPPE, S. SCHENKLER, H. D. ULRICH, T. U. MAYER, and S. JENTSCH. 1999. A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. Cell **96**:635–644.

KUBOTA, H., Y. SAKAKI, and T. ITO. 2000. GI domain-mediated association of the eukaryotic initiation factor 2α kinase GCN2 with its activator GCN1 is required for general amino acid control in budding yeast. J. Biol. Chem. **275**:20243–20246.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, and M. NEI. 2001. MEGA2: molecular evolutionary genetics analysis software. Arizona State University, Tempe, Arizona.

LIM, S., C. SALA, J. YOON, S. PARK, S. KURODA, M. SHENG, and E. KIM. 2001. Sharpin, a novel postsynaptic density protein that directly interacts with the Shank family of proteins. Mol. Cell. Neurosci. **17**:385–397.

LORICK, K. L., J. P. JENSEN, S. FANG, A. M. ONG, S. HATAKEYAMA, and A. M. WEISSMAN. 1999. RING fingers mediate ubiquitin-conjugating enzyme (E2)-dependent ubiquitination. Proc. Natl. Acad. Sci. USA **96**:11364–11369.

MARCOTTE, E. M., M. PELLEGRINI, H. L. NG, D. W. RICE, T. O. YEATES, and D. EISENBERG. 1999. Detecting protein function and protein-protein interactions from genome sequences. Science **285**:751–753.

MARTINEZ-NOEL, G., R. NIEDENTHAL, T. TAMURA, and K. HARBERS. 1999. A family of structurally related RING finger proteins interacts specifically with the ubiquitin-conjugating enzyme UbcM4. FEBS Lett. **454**:257–261.

MORETT, E., and P. BORK. 1999. A novel transactivation domain in parkin. Trends Biochem. Sci. **24**:229–231.

MOYNIHAN, T. P., H. C. ARDLEY, U. NUBER, S. A. ROSE, P. F. JONES, A. F. MARKHAM, M. SCHEFFNER, and P. A. ROBINSON. 1999. The Ubiquitin-conjugating enzymes UbcH7 and UbcH8 interact with RING finger/IBR motif-containing domains in HHARI and H7-AP1. J. Biol. Chem. **274**:30963–30968.

MURALIDHAR, M. G., and J. B. THOMAS. 1993. The *Drosophila bendless* gene encodes a neural protein related to ubiquitin-conjugating enzymes. Neuron **11**:253–266.

NEI, M., and S. KUMAR. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

NICHOLAS, K. B., and H. B. NICHOLAS JR. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignments. Distributed by the author (www.cris.com/~ketchup/genedoc.shtml).

NIWA, J., S. ISHIGAKI, M. DOYU, T. SUZUKI, K. TANAKA, and G. SOBUE. 2001. A novel centrosomal RING-finger protein, Dorfin, mediates ubiquitin ligase activity. Biochem. Biophys. Res. Commun. **281**:706–713.

NOMURA, N., T. NAGASE, N. MIYAJIMA, T. SAZUKA, A. TANAKA, S. SATO, N. SEKI, Y. KAWARABAYASI, K. ISHIKAWA, and S. TABATA. 1994. Prediction of the coding sequences of unidentified human genes. II. The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by analysis of cDNA clones from human cell line KG-1. DNA Res. **1**: 223–229.

OH, C. E., R. MCMAHON, S. BENZER, and M. A. TANOUYE. 1994. *bendless*, a *Drosophila* gene affecting neuronal connectivity, encodes a ubiquitin-conjugating enzyme homolog. J. Neurosci. **14**:3166–3179.

OKAMURA, K., Y. HAGIWARA-TAKEUCHI, T. LI, T. H. VU, M. HIRAI, M. HATTORI, Y. SAKAKI, A. R. HOFFMAN, and T. ITO. 2000. Comparative genome analysis of the mouse imprinted

gene *Impact* and its nonimprinted human homolog *IMPACT:* toward the structural basis for species-specific imprinting. Genome Res. **10**:1878–1889.

PAGE, R. D. M. 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12**:357–358.

PÁRRAGA, M., and J. DEL MAZO. 2000. XYbp, a novel RING-finger protein, is a component of the XY body of spermatocytes and centrosomes. Mech. Dev. **90**:95–101.

POLYMEROPOULOS, M. H. 2000. Genetics of Parkinson's disease. Ann. N Y Acad. Sci. **920**:28–32.

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

SAURIN, A. J., K. L. BORDEN, M. N. BODDY, and P. S. FREEMONT. 1996. Does this have a familiar RING? Trends Biochem. Sci. **21**:208–214.

SHIMURA, H., N. HATTORI, S. I. KUBO et al. (11 co-authors). 2000. Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat. Genet. **25**:302–305.

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

TOKUNAGA, C., S. KURODA, K. TATEMATSU, N. NAKAGAWA, Y. ONO, and U. KIKKAWA. 1998. Molecular cloning and characterization of a novel protein kinase C-interacting protein with structural motifs related to RBCC family proteins. Biochem. Biophys. Res. Commun. **244**:353–359.

TYERS, M., and P. JORGENSEN. 2000. Proteolysis and the cell cycle: with this RING I do thee destroy. Curr. Opin. Genet. Dev. **10**:54–64.

TYERS, M., and A. R. WILLEMS. 1999. One ring to rule a superfamily of E3 ubiquitin ligases. Science **284**:601–604.

UETZ, P., L. GIOT, G. CAGNEY et al. (20 co-authors). 2000. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature **403**:623–627.

VALENTE, E. M., A. R. BENTIVOGLIO, P. H. DIXON, A. FERRARIS, T. IALONGO, M. FRONTALI, A. ALBANESE, and N. W. WOOD. 2001. Localization of a novel locus for autosomal recessive early-onset parkinsonism, PARK6, on human chromosome 1p35-p36. Am. J. Hum. Genet. **68**:895–900.

VAN DER REIJDEN, B. A., C. A. J. ERPELINCK-VERSCHEREN, B. LÖWENBERG, and J. H. JANSEN. 1999. TRIADs: a new class of proteins with a novel cysteine-rich signature. Prot. Sci. **8**:1557–1561.

YU, H., J. M. PETERS, R. W. KING, A. M. PAGE, P. HIETER, and M. W. KIRSCHNER. 1998. Identification of a cullin homology region in a subunit of the anaphase-promoting complex. Science **279**:1219–1222.

ZHANG, Y., J. GAO, K. K. K. CHUNG, H. CHUNG, V. L. DAWSON, and T. M. DAWSON. 2000. Parkin functions as an E2-dependent ubiquitin-protein ligase and promotes the degradation of the synaptic vesicle-associated protein, CDCrel-1. Proc. Natl. Acad. Sci. USA **97**:13354–13359.

ZHAO, X., C. T. SMARTT, J. F. HILLYER, and B. M. CHRISTENSEN. 2000. A novel member of the RING-finger gene family associated with reproductive tissues of the mosquito, *Aedes aegypti*. Insect Mol. Biol. **9**:301–308.

DIETHARD TAUTZ, reviewing editor

Accepted May 6, 2002