

## **The Parkinson disease gene *LRRK2*: evolutionary and structural insights**

Ignacio Marín

Departamento de Genética, Universidad de Valencia, Burjassot, Spain.

Full address:

Ignacio Marín  
Departamento de Genética  
Universidad de Valencia  
Calle Doctor Moliner, 50  
Burjassot 46100 (Valencia)  
Spain

E-mail: [ignacio.marin@uv.es](mailto:ignacio.marin@uv.es)  
Tel.: 34 – 963544504  
Fax: 34 – 963543029

## Abstract

Mutations in the human *Leucine-rich repeat kinase 2 (LRRK2)* gene are associated with both familial and sporadic Parkinson disease (PD). *LRRK2* belongs to a gene family known as Roco. Roco genes encode for large proteins with several protein domains. Particularly, all Roco proteins have a characteristic GTPase domain, named Roc, plus a domain of unknown function called COR. In addition, *LRRK2* and several other Roco proteins also contain a protein kinase domain. In this study, I use a combination of phylogenetic and structural analyses of the COR, Roc and kinase domains present in Roco proteins to describe the origin and evolutionary history of *LRRK2*. Phylogenetic analyses using the ROC domain demonstrate that *LRRK2* emerged from a duplication that occurred after the protostome-deuterostome split. The duplication was followed by the acquisition by *LRRK2* proteins of a specific type of N-terminal repeat, described here for the first time. This repeat is absent in the proteins encoded by the paralogs of *LRRK2*, called *LRRK1* or in protostome LRRK proteins. These results suggest that *Drosophila* or *Caenorhabditis* *LRRK* genes may not be good models to understand human *LRRK2* function. Genes in the slime mold *Dictyostelium discoideum* with structures very similar to those found in animal *LRRK* genes, including the protein kinase domain, have been described. However, phylogenetic analyses suggest this structural similarity is due to independent acquisitions of distantly related protein kinase domains. Finally, I confirm in an extensive sequence analysis that the Roc GTPase domain is related but still substantially different from small GTPases such as Rab, Ras or Rho. Modelling based on known kinase structures suggests that mutations in *LRRK2* that cause familial PD may alter the local three-dimensional folding of the *LRRK2* protein without affecting its overall structure.

## Introduction

Parkinson disease (PD) is the second most common neurodegenerative disease and therefore the characterization of its causes and the discovery of palliative treatments or, if possible, ways of curing the disease are one of the main battlefields in modern medicine. Although most PD cases are sporadic, a small percentage of them are due to genetic causes and the last years have witnessed the discovery of several genes which mutations strongly contribute to the generation of PD (see recent reviews by Abou-Sleiman, Muqit and Wood 2006; Farrer 2006). In October 2004, two studies demonstrated an involvement of mutations in the *LRRK2* gene in autosomal dominant familial Parkinson disease (Paisán-Ruiz et al. 2004; Zimprich et al. 2004). Those seminal findings have been shortly followed by a large number of additional studies demonstrating that *LRRK2* mutations are not only often involved in familial PD (reviewed in Taylor, Mata and Farrer 2006) but also in the most common, sporadic, form of the disease (Gilks et al. 2005; Skipper et al. 2005; Mata et al. 2006). *LRRK2* mutations have been estimated to be involved in up to 13% of familial and up to 3% of sporadic PD cases (Berg et al. 2005; Taylor, Mata and Farrer 2006).

Before *LRRK2* was related to PD, a few researchers became interested in this gene because of its obvious relationship with several *Dyctiostelium discoideum* genes involved in cytokinesis, cell polarity and chemotaxis (Bosgraaf et al. 2002; Goldberg et al. 2002; Abe et al. 2003; Abysalh, Kuchnicki and Larochele 2003). This led two of them, Bosgraaf and Van Haastert (2003), to describe the Roco family, that includes all these *D. discoideum* genes plus genes found in prokaryotes, plants and animals. One of the animal genes, which they called “human Roco2”, corresponds to the *LRRK2* gene. All Roco family genes encode long proteins with two characteristic domains. The first, called Roc, is similar to small GTPases of the Ras superfamily. The second is a domain of unknown function that was named COR (Bosgraaf

and Van Haastert 2003). In addition, other domains appear in several of the Roco proteins. The two most common are typical leucine-rich repeats (LRRs), located N-terminally respect to the Roc domain, and protein kinase domains. Both of them are present in LRRK2 proteins, and mutations in the LRRs, Roc, COR or protein kinase domains have been found in PD-affected individuals (reviewed in Taylor, Mata and Farrer 2006).

The *LRRK2* gene is expressed in multiple tissues and in multiple brain regions in humans and rodents (Zimprich et al. 2004; Paisán-Ruiz et al. 2004; Galter et al. 2006; Giasson et al. 2006; Melrose et al. 2006; Simón-Sánchez et al. 2006). Its cellular functions are so far largely unknown. The finding in LRRK2 of a Ras-like GTPase domain plus a protein kinase domain quite similar in sequence to Raf suggested an obvious paralellism with the beginning of the Ras signal transduction pathway: the kinase domain of LRRK2 might be activated by a GDP to GTP transition in its GTPase domain. There is some evidence that this is actually the case for the protein encoded by the paralog of *LRRK2*, *LRRK1* (Korr et al. 2006). It has been found that *LRRK2* missense mutations associated to dominant PD generate proteins with increased kinase activity and, in cell culture assays, are able to induce the generation of inclusion bodies that lead to cell death (West et al. 2005; Gloeckner et al. 2006; Greggio et al. 2006). On the contrary, mutations that eliminate kinase activity inhibit the formation of inclusion bodies in cell cultures (Greggio et al. 2006). These results strongly suggest that the dominant effect of these mutations are due to hyperactivity of the resulting proteins and not due to loss of function and haploinsufficiency. LRRK2 protein has been found to interact with Parkin in cell culture assays (Smith et al. 2005). Parkin belongs to the RBR family of ubiquitin ligases (Marín and Ferrús 2002; Marín et al. 2004) and mutations in the *parkin* gene are a well-known cause of familial PD (reviewed in Abou-Sleiman, Muqit and Wood 2006; Farrer 2006). These results, together with the fact that Lewy bodies and other proteinaceous

inclusions are found in individuals affected by LRRK2 mutations (Zimprich et al. 2004; Wszolek et al. 2004; Giasson et al. 2006) suggest a potential involvement of LRRK2 in regulation of ubiquitin metabolism. So far, no *in vivo* animal models for *LRRK2* have been described.

My group has been recently focused on tracing the evolutionary history of Parkinson disease genes in order to provide novel hints about their cellular functions (Marín and Ferrús 2002; Marín et al. 2004; Lucas, Arnau and Marín 2006). In this study, I describe a comprehensive set of comparative genomics and structural analyses devised to determine the origin and evolutionary history of the *LRRK2* genes. The goal is to provide a framework in which to base further experimental approaches and, especially, to choose appropriate animal models in which to explore the functions of this gene.

## **Methods**

### ***Phylogenetic analyses***

BLASTP and TBLASTN searches were performed against the National Center for Biotechnology Information (NCBI) databases (<http://www.ncbi.nlm.nih.gov/>) using the COR, Roc or kinase domains of several Roco domain proteins as queries. For the COR domain, I pursued the searches until results became saturated, thus detecting all significant matches. However, this strategy could not be used for either GTPase or kinase domain sequences, because the number of these sequences in the databases is too large. Therefore, for the Roc domain, a large number of the most significant matches (that almost exclusively belonged to the Rab family of small GTPases) were obtained and then representative sequences of the

Ras, Ran, Rho and Arf families were manually added. These last sequences were obtained from the Smart (<http://smart.embl-heidelberg.de/>; domains SM00176 and SM00173), Pfam (<http://www.sanger.ac.uk/Software/Pfam/>; domains PF00071, PF00025) and Conserved Domains (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cdd>; domains cd00154, cd00157) databases. Similarly, for the kinase domain I took a large number of sequences with high level of similarity to the kinase domains of Roco proteins. Almost all of them belonged to the TKL class, to which Roco domains have been assigned (see Manning et al. 2002; Goldberg et al. 2006). Then, I manually added several other human kinases, selected according to their proximity to LRRK2 in the trees obtained by Manning et al. (2002) for the human kinome. These human kinase sequences were obtained from the Protein Kinase Resource (<http://www.kinasenet.org/pkr/>).

The protein sequences found were then aligned using ClustalX version 1.83 (Thompson et al. 1997) and preliminary phylogenetic trees using the Neighbor-Joining (NJ; Saitou and Nei 1987) routine available in ClustalX 1.83 were obtained. Those trees were used to detect duplicates and partial sequences, which were eliminated. After this process, the corrected databases contained 86 (COR domain), 370 (Roc domain) and 400 (kinase domain) sequences. From them, I generated the final multiple-protein alignments shown below, again using first ClustalX 1.83 and then Genedoc 2.6 (Nicholas, Nicholas and Deerfield 1997) to manually correct the alignments. Phylogenetic trees were then obtained both by the NJ and the maximum-parsimony (MP) methods, using the routines available in MEGA 3.1 (Kumar, Tamura and Nei 2004) and PAUP\*, beta 10 version (Swofford 2003), respectively. For NJ, sites with gaps were included and Kimura's correction was used, whereas for MP, the parameters were as follows: (1) all sites included, (2) randomly-generated trees used as seeds,

(3) maximum number of tied trees saved equal to 20; (4) heuristic search using the subtree pruning-regrafting algorithm. Support for the topologies obtained with those two methods was determined using the bootstrap routines also available in MEGA 3.1 and PAUP\*. 1000 replicates were performed for both NJ and MP bootstrap analyses. Phylogenetic trees were depicted using the tree editor of MEGA 3.1.

### ***Structural analyses***

Domain searches were performed against the Pfam, Smart and Conserved Domain databases, already cited. Motif searches with the human and sea urchin LRRK2-specific repeats were performed using PRATT 2.1 (Jonassen, Collins and Higgins 1995; <http://www.expasy.org/tools/pratt/>; minimum match: 25%) to generate sequence patterns and ScanProsite (Gattiker et al. 2002; <http://www.expasy.org/cgi-bin/scanprosite>) to determine whether those patterns were present in other proteins in the PROSITE database. Three-dimensional structures were predicted with Swiss-Model (Peitsch 1996; online at <http://swissmodel.expasy.org/>) using the crystal structures of either Ras superfamily GTPase or kinase proteins as templates (Protein Data Bank codes 2ew1A and 2bmeA-D for the GTPase structures and codes 1uwjA-B, 1fotA, 1k9aA and 2fh9A for the kinase structures). Swiss-Pdb viewer 3.7 (Guex and Peitsch 1997) was used to generate the three-dimensional images shown below.

## Results

### *COR domain comparisons allow to trace the origin of LRRK2*

The COR domain is common to all Roco proteins and sufficiently large as to provide enough information to characterize the relationships among the main groups of the family. Figure 1 shows the results obtained using both neighbor-joining and maximum parsimony phylogenetic reconstructions. Mammalian LRRK1 and LRRK2 sequences appear in a monophyletic group together with several sequences obtained from invertebrate species and clearly separated from the rest of Roco proteins. All protostomes have a single *LRRK* gene. This result, together with the finding of genes very closely related to both *LRRK1* and *LRRK2* in the sea urchin *Strongylocentrotus purpuratus* strongly suggests that *LRRK2* originated by a gene duplication shortly after the protostome-deuterostome split. In Figure 1, it can be observed that the closest relatives to animal *LRRK* genes are the large set of *D. discoideum* sequences described by Bosgraaf and Van Haastert (2003). However, this association is not supported by significant bootstrap values.

The rest of sequences included in this tree mostly correspond to those already detected by Bosgraaf and Van Haastert (2003), which included the animal *MFHAS1* (also known as *MASL1*) and *DAPK1* genes (Deiss et al. 1995; Sakabe et al. 1999) plus a few genes from both eubacteria and archaea and from plants. However, a couple of novel *MFHAS1*-like sequences were found in two species, *Danio rerio* and *Gallus gallus* (see Figure 1). Finally, it is significant that the classification in three groups, based on structural similarity, proposed by Bosgraaf and Van Haastert (2003) is not supported by this phylogenetic analysis.

### ***Singularity of the Roc GTPase domain***

Bosgraaf and Van Haastert (2003) performed a relatively limited phylogenetic analyses including 21 Roc GTPase domains plus 34 domains belonging to small GTPases of the Ras, Rho, Rab, Ran and Arf families. They found that Roc domains appeared as a monophyletic group, separated from the rest of GTPases, although with low bootstrap support. However, BLAST searches using Roc domains as queries always detect Rab GTPases as having the highest similarity scores (not shown). This result might be significant, because it would suggest a potential functional similarity between Rab GTPases and Roco proteins. Therefore, I decided to perform an extensive analysis to determine whether Bosgraaf and Van Haastert (2003) results were due to incomplete sampling of Rab proteins. Figure 2 shows the results of an analysis with 62 Roc domains plus other 308 small GTPase domains, including all the Rab sequences with the highest similarity to Roco family sequences. Although bootstrap values are quite low, the Roco family sequences again appear as a monophyletic group separated from the rest of GTPases. Thus, these results fully confirm Bosgraaf and Van Haastert (2003) previous findings, while the particularly close similarity to Rab family proteins suggested by BLAST analyses is not supported. On closer inspection, it can be determined that the BLAST results are mainly due to differences in domain size: Roc domain sequences are slightly more similar in size to the Rab sequences than to the sequences of the rest of families.

### ***Protein kinase domain analyses do not support a monophyletic origin for animal and D. discoideum Roco genes***

Two clearly distinct groups of animal Roco genes, *LRRK* genes and *DAPK1* genes, contain kinase domains. This domains are however very different. Kinase domains in *LRRK* genes can be classified as belonging to the TKL (tyrosine kinase-like) group, while the

*DAPK1* kinase domain can be included in the CAMK (calcium/calmodulin-dependent kinase) group (Manning et al. 2002). The fact that kinase domains have been coopted at least twice in a relatively short period of time is striking. Interestingly, apart from these animal genes, the only ones that also have a kinase domain are several *Dyctiostelium discoideum* genes. The kinase domains of these genes are clearly related to those in animal LRRKs. However, the question of whether this similarity is due to common ancestry or to independent coptions of closely related kinase domains has never been tackled. Figure 3 shows the phylogenetic trees obtained for the *LRRK* and *D. dictyostelium* kinase domains plus their closest relatives in BLAST searches and a set of selected kinases of the TKL group. Notably, the animal and *Dyctiostelium* sequences appear as two separated groups in distant positions in these trees. This result suggest that independent coptions for related kinase domains have occurred in animal and *Dyctiostelium* genes. The low bootstrap support for the inner branches of the tree preclude to establish from which specific type of kinases these domains were acquired.

### ***Structural analyses of LRRK proteins and implications for research in model animals***

There is some confusion in the literature respect to the structures of LRRK proteins. Most authors, depict human LRRK2 protein as having, going from the N- to the C-terminal end, several LRRs (four according to Bosgraaf and Van Haastert [2003], 13 according to Guo, Wang and Chen [2006]), single Roc, COR and kinase domains and, finally one (Mata et al. 2006), two (Bosgraaf and Van Haastert 2003) or even seven (Guo, Wang and Chen 2006) WD40 repeats. Moreover, some authors indicate the presence of N-terminal ankyrin repeats (e. g. West et al. 2005; Mata et al. 2006). In fact, structural analyses using the Pfam, SMART and Conserved domain (NCBI) databases support the existence in human LRRK2 proteins of several LRRs (seven according to SMART, eight according to Pfam) and obviously also of

the Roc and kinase domains, but support for WD40 repeats is weak (a single WD40 repeat being in position 2231 to 2276 is predicted by SMART). No ankyrin repeats are ever predicted. The finding of evolutionary distant *LRRK1* and *LRRK2* relatives allows to check whether the canonical structures detected in the human proteins are evolutionarily conserved. Figure 4 shows the details for the structures of LRRK proteins in invertebrates, both protostomes and deuterostomes, compared with the human proteins, according to Pfam and my own observations. Two results are noteworthy. First, the presence in most LRRK proteins of N-terminal ankyrin repeats. The fact that both sea urchin LRRK1 and LRRK2 proteins contain ankyrin repeats suggest that the loss of these repeats, as seen in mammalian LRRK2s (Figure 4, top), is quite recent. Second, I have detected that the whole N-terminal region of LRRK2 proteins in both sea urchins and humans is characterized by having a total of fourteen evolutionary conserved repeats. They are specific of this type of proteins, not appearing either in LRRK1 or in any other proteins available in databases such as PROSITE. The sequences of these repeats, located along the first 660 (human) to 850 (sea urchin) amino acids of LRRK2 proteins are detailed in Figure 5.

Several authors have already shown that the Roc and kinase domains of LRRK2 proteins are sufficiently similar to other proteins for which crystal structures are available as to allow a prediction of their three-dimensional structures (e.g. Guo, Wang and Chen 2006; Mata et al. 2006; Tan et al. 2006). An interesting point hitherto unexplored is whether known mutations that affect those domains and may induce PD are able to substantially change their spatial structures. In Figure 6, a model for the wild-type kinase domain of LRRK2 (Figure 6B) is shown that closely resembles the structure of related kinases, such as human B-Raf (Figure 6A). However, the kinase G2019S and I2020T mutations, both known to be associated to PD, change the local structure of the kinase domain (see asterisks in Figures 6C,

6D) but without affecting its overall folding. On the contrary, no obvious change was observed when a PD-associated mutation in the Roc GTPase domain (R1441C) was modelled (not shown).

## **Discussion**

This study is focused on describing the evolution and structural characteristics of Roco family genes, with emphasis in the PD gene *LRRK2*. To trace the general evolutionary history of this group of genes, analyses of the COR domain have been performed (see Figure 1). Roco family genes are present in prokaryotes, both eubacteria (cyanobacteria, proteobacteria, planctomycetes, chlorobi) and archaea. They have been also detected in a few plant species, in the slime mold *D. discoideum* (in which a large amplification of this type of genes has occurred) and in animals. This patchy phylogenetic distribution of the Roco family is difficult to understand, but the two most likely explanations are: 1) a very ancient origin previous to the origin of eukaryotes and 2) an origin in early eukaryotic history followed by horizontal transmission to prokaryotic species. In both cases, losses in multiple lineages must be hypothesized. More complex evolutionary histories, involving several horizontal transfer events, cannot be excluded at present.

If we focus on understanding human *LRRK2* gene function, it is crucial to determine the origin of the gene. Data presented above shows that genes significantly related to *LRRK2* have a narrow phylogenetic range. First, the LRR-GTPase-kinase structure typical of the proteins encoded by animal *LRRK* genes originated recently. Significantly, the structural similarity of *Dictyostelium* Roco genes and *LRRK* genes pinpointed by Bosgraaf and Van Haastert (2003) is likely a convergent feature, due to two independent cooptions of relatively

similar kinase domains (Figure 3). Second, vertebrate-specific amplification of this family has occurred: protostomes have only 1-2 Roco genes while up to five can be found in vertebrates. COR domain analyses have shown that *LRRK2* emerged by gene duplication quite recently, after the protostome/deuterostome split (Figure 1). Moreover, protostome LRRK proteins are structurally much more similar to deuterostome LRRK1 proteins than to deuterostome LRRK2 proteins (Figure 4). These results have obvious experimental implications: analysis of the *LRRK* genes of commonly used protostome model species such as *Drosophila* or *Caenorhabditis* may not be appropriate to understand the cellular functions of human *LRRK2*. All these data together mean that the best model species in which to explore *LRRK2* function are deuterostomes (e. g. echinoderms and chordates), the only groups in which *bona fide* orthologs of human *LRRK2* have been found.

As we have shown in several previous works, significant clues about the roles of genes involved in human diseases can be obtained by understanding their phylogenetic context and the structural features of their products (e. g. Marín and Ferrús 2002; Marco et al. 2004; Marín et al. 2004; Lucas, Arnau and Marín 2006). Data from the Roc (Figure 2) and kinase domains (Figure 3) show that they are quite different from any other GTPases or kinases found in the databases, warning against simplified views of LRRK2 proteins as an ensemble of a Ras-like GTPase plus a Raf-like kinase. Actually, according to their sequences neither Ras nor Raf are closely related to the corresponding domains in LRRK2 (see again Figures 2 and 3). The complex structures of Roco proteins, and most especially of those encoded by *LRRK2* genes (Figure 4), are difficult to reconcile with what we know about related GTPase or kinase families, suggesting that LRRK2 proteins are performing genuinely novel functions, specific of deuterostome species.

Many missense, likely hyperactivity/gain-of-function *LRRK2* mutations associated to PD have been described, and most of them affect the obvious domains of this protein (LRRs, Roc, COR and kinase domain). In an excellent review; Mata et al. (2006) discuss the potential structural implications of those mutations. Two additional data are derived from this study. First, so far no known mutations related to PD affect the LRRK2-specific repeats, described here for the first time. Second, the two mutations in the kinase domain most likely generate significant local changes of the three-dimensional structure of that domain but without affecting their overall folding (Figure 6). These results are compatible with the mutant kinase domains being active, as shown in recent experiments (West et al. 2005; Gloeckner et al. 2006).

### **Acknowledgements**

This research was supported by grant SAF2003-09506 (Ministerio de Educación y Ciencia, Spain).

## Literature cited

Abe, T., J. Langenick, and J. G. Williams. 2003. Rapid generation of gene disruption constructs by in vitro transposition and identification of a *Dictyostelium* protein kinase that regulates its rate of growth and development. *Nucleic Acids Res.* **31**:e107.

Abysalh, J.C., L. L. Kuchnicki LL, and D. A. Larochelle. 2003. The identification of *pats1*, a novel gene locus required for cytokinesis in *Dictyostelium discoideum*. *Mol. Biol. Cell.* **14**:14-25.

Abou-Sleiman, P.M., M. M. K. Muqit, and N. W. Wood. 2006. Expanding insights of mitochondrial dysfunction in Parkinson's disease. *Nat. Rev. Neurosci.* **7**:207-219.

Berg, D., Schweitzer, K., Leitner, P., et al. (10 authors). 2005. Type and frequency of mutations in the LRRK2 gene in familial and sporadic Parkinson's disease. *Brain* **128**:3000-3011.

Bosgraaf, L., and P. J. M. Van Haastert. 2003. Roc, a Ras/GTPase domain in complex proteins. *Biochem. Biophys. Acta* **1643**:5-10.

Bosgraaf, L., H. Russcher, J. L. Smith, D. Wessels, D. R. Soll, and P. J. Van Haastert. 2002. A novel cGMP signalling pathway mediating myosin phosphorylation and chemotaxis in *Dictyostelium*. *EMBO J.* **21**:4560-4570.

Deiss, L. P., E. Feinstein, H. Berissi, O. Cohen, and A. Kimchi. 1995. Identification of a novel serine/threonine kinase and a novel 15-kD protein as potential mediators of the gamma interferon-induced cell death. *Genes Dev.* **9**:15-30.

Farrer, M. J. 2006. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat. Rev. Genet.* **7**:306-18.

Galter, D., M. Westerlund, A. Carmine, E. Lindqvist, O. Sydow, and L. Olson. 2006. LRRK2 expression linked to dopamine-innervated areas. *Ann. Neurol.* **59**:714-719.

Gattiker, A., E. Gasteiger, and A. Bairoch. 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics*. **1**:107-108.

Guex, N. and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**:2714-2723.

Giasson, B. I., J. P. Covy, N. M. Bonini, H. I. Hurtig, M. J. Farrer, J. Q. Trojanowski, and V. M. Van Deerlin. 2006. Biochemical and pathological characterization of Lrrk2. *Ann. Neurol.* **59**:315-322.

Gilks, W. P., P. M. Abou-Sleiman, S. Gandhi, et al. (15 authors). 2005. A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet*. **365**:415-416.

Goldberg, J.M., L. Bosgraaf, P. J. Van Haastert, and J. L. Smith. 2002 Identification of four candidate cGMP targets in *Dictyostelium*. *Proc. Natl. Acad. Sci. U S A*. **99**:6749-6754.

Goldberg, J.M., G. Manning, A. Liu, P. Fey, K. E. Pilcher, Y. Xu, and J. L. Smith. 2006. The *Dictyostelium* kinome -- analysis of the protein kinases from a simple model organism. *PLoS Genet*. **2**:e38.

Gloeckner C. J., N. Kinkl, A. Schumacher, R. J. Braun, E. O'Neill, T. Meitinger, W. Kolch, H. Prokisch, and M. Ueffing. 2006 The Parkinson disease causing *LRRK2* mutation I2020T is associated with increased kinase activity. *Hum. Mol. Genet.* **15**:223-232.

Greggio E., S. Jain, A. Kingsbury, et al (18 authors). 2006. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol. Dis.* (in press)

Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**:2714-2723.

Guo, L., W. Wang, and S. G. Chen. 2006 Leucine-rich repeat kinase 2: Relevance to Parkinson's disease. *Int. J. Biochem. Cell Biol.* **38**:1469-1475.

- Jonassen, I., J. F. Collins, and D. G. Higgins. 1995. Finding flexible patterns in unaligned protein sequences. *Protein Science* **4**:1587-1595.
- Korr, D., L. Toschi, P. Donner, H. D. Pohlenz, B. Kreft, and B. Weiss. 2006. LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. *Cell Signal* **18**:910-20.
- Kumar, S., K. Tamura, M. Nei. 2004. MEGA3: Integrated Software for Molecular Evolutionary Genetics Analysis and Sequence Alignment. *Briefings in Bioinformatics* **5**:150-163.
- Lucas, J. I., V. Arnau, and I. Marín. 2006. Comparative genomics and protein domain graph analyses link ubiquitination and RNA metabolism. *J. Mol. Biol.* **357**:9-17
- Manning, G., D. B. Whyte, R. Martinez, T. Hunter, and S. Sudarsanam. 2002. The protein kinase complement of the human genome. *Science* **298**:1912-1934.
- Marín, I., and A. Ferrús. 2002. Comparative genomics of the RBR family, including the Parkinson's disease-related gene *parkin* and the genes of the ariadne subfamily. *Mol. Biol. Evol.* **19**:2039-2050.
- Marín, I., J. I. Lucas, A. C. Gradilla, and A. Ferrús. 2004. Parkin and relatives: the RBR family of ubiquitin ligases. *Physiol. Genomics* **17**:253–263
- Marco, A., A. Cuesta, L. Pedrola, F. Palau, I. Marín. 2004. Evolutionary and structural analyses of *GDAP1*, involved in Charcot-Marie-Tooth disease, characterize a novel class of glutathione transferase-related genes. *Mol. Biol. Evol.* **21**:176-187.
- Mata, I. F., W. J. Wedemeyer, M. J. Farrer, J. P. Taylor, and K. A. Gallo. 2006. LRRK2 in Parkinson's disease: protein domains and functional insights. *Trends Neurosci.* **29**:286-293.
- Melrose, H., S. Lincoln, G. Tyndall, D. Dickson, and M. Farrer. 2006. Anatomical localization of leucine-rich repeat kinase 2 in mouse brain. *Neuroscience* **139**:791-794.

- Nicholas, K. B., H. B. Nicholas Jr., and D. W. Deerfield. 1997 GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS* **4**:14
- Paisán-Ruiz, C., S. Jain, E. W. Evans, et al. (21 authors). 2004. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44**:595-600.
- Peitsch, M. C. 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **24**:274–279.
- Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- Sakabe, T., T. Shinomiya, T. Mori, Y. Ariyama, Y. Fukuda, T. Fujiwara, Y. Nakamura, and J. Inazawa. 1999. Identification of a novel gene, *MASL1*, within an amplicon at 8p23.1 detected in malignant fibrous histiocytomas by comparative genomic hybridization. *Cancer Res.* **59**:511-515.
- Simon-Sanchez, J., V. Herranz-Perez, F. Olucha-Bordonau, and J. Perez-Tur. 2006. LRRK2 is expressed in areas affected by Parkinson's disease in the adult mouse brain. *Eur. J. Neurosci.* **23**:659-66.
- Skipper, L., Y. Li, C. Bonnard et al. (11 authors). 2005. Comprehensive evaluation of common genetic variation within LRRK2 reveals evidence for association with sporadic Parkinson's disease. *Hum. Mol. Genet.* **14**:3549-3556.
- Smith, W. W., Z. Pei, H. Jiang, D. J. Moore, Y. Liang, A. B. West, V. L. Dawson, T. M. Dawson, and C. A. Ross. 2005. Leucine-rich repeat kinase 2 (LRRK2) interacts with parkin, and mutant LRRK2 induces neuronal degeneration. *Proc. Natl. Acad. Sci. USA* **102**:18676-18681.
- Swofford, D. L. 2003. PAUP\*. Phylogenetic Analysis Using Parsimony (\*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tan, E. K., L. Skipper, E. Chua, M. C. Wong, R. Pavanni, C. Bonnard, P. Kolatkar, and J. J. Liu. 2006. Analysis of 14 LRRK2 mutations in Parkinson's plus syndromes and late-onset Parkinson's disease. *Mov. Disord.* (in press)

Taylor, J. P., I. F. Mata, and M. J. Farrer. 2006. LRRK2: a common pathway for parkinsonism, pathogenesis and prevention?. *Trends Mol. Med.* **12**:76-82.

Tchieu, J. H., F. Fana, J. L. Fink, et al. (12 authors). 2003. The PlantsP and PlantsT Functional Genomics Databases. *Nucleic Acids Res.* **31**:342-344.

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.

West, A. B., D. J. Moore, S. Biskup, A. Bugayenko, W. W. Smith, C. A. Ross, V. L. Dawson, and T. M. Dawson. 2005. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl. Acad. Sci. U S A.* **102**:16842-16847.

Wszolek, Z. K., R. F. Pfeiffer, Y. Tsuboi, et al. (13 authors). 2004. Autosomal dominant parkinsonism associated with variable synuclein and tau pathology. *Neurology.* **62**:1619-1622.

Zimprich, A., S. Biskup, P. Leitner, et al. (22 authors). 2004. Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **44**:601-607.

## FIGURE LEGENDS

**Figure 1.** Phylogenetic trees obtained with the COR domain sequences. The main groups of vertebrate genes (*LRRK1*, *LRRK2*, *MFHAS1*, *MFHAS1*-like and *DAPK1*) are detailed. Numbers refer to bootstrap support for both methods of phylogenetic reconstruction (in the order NJ/MP). Values are shown only when both NJ and MP methods supported the same branch and at least one of the two bootstrap values was higher than 50%. Asterisks indicate branches in which both NJ and MP bootstrap values were  $\geq 95\%$ . A few significant bootstrap values for very short terminal branches in the *LRRK1* and *DAPK1* animal groups have been omitted for simplicity.

**Figure 2.** Tree showing the relationships of the main families of small GTPases of the Ras superfamily and the Roc GTPase domain found in Roco proteins. Bootstrap support shown as in Figure 1. Numbers in brackets refer to the number of sequences in the branch.

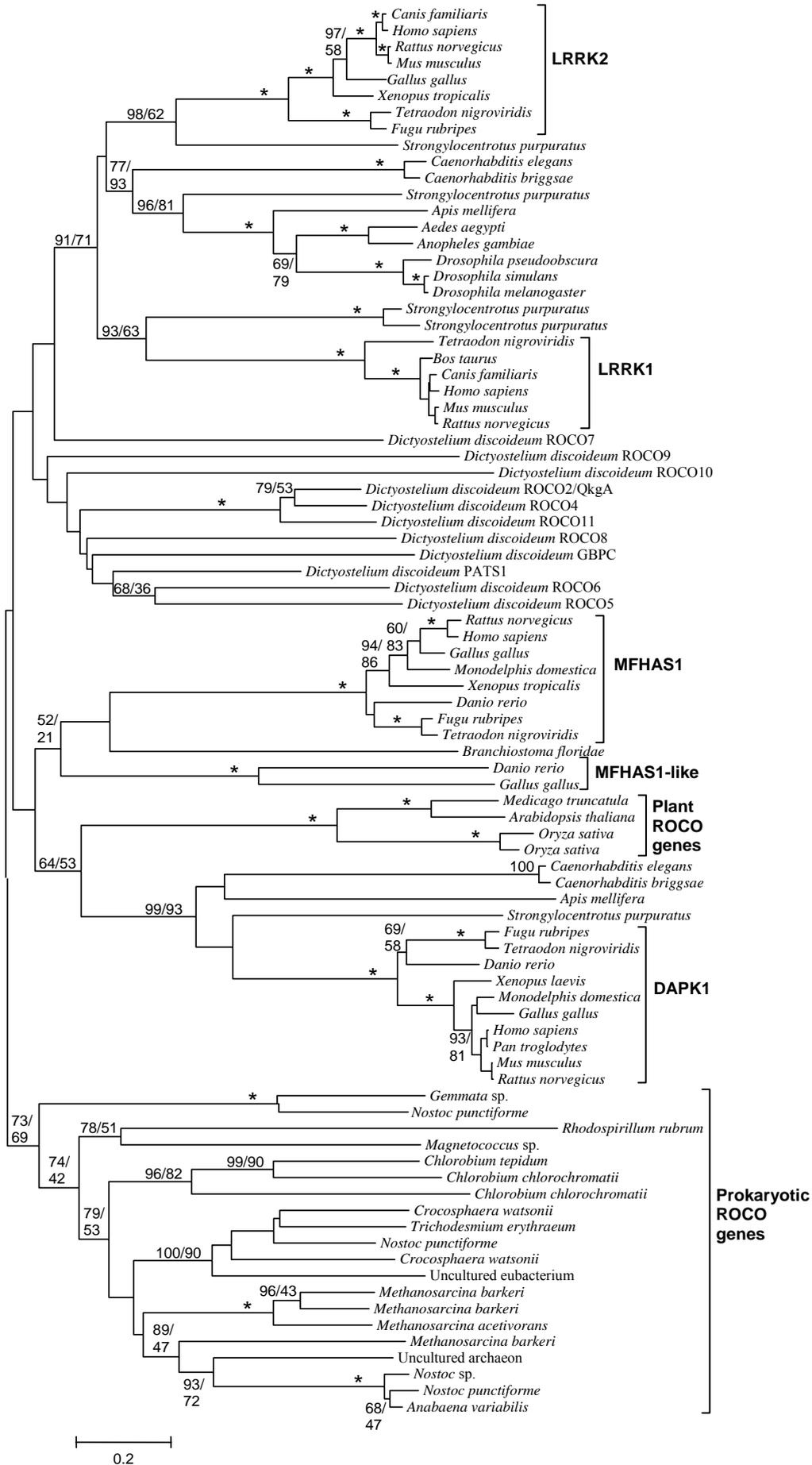
**Figure 3.** Phylogenetic trees for kinase domains. Bootstrap values and number of sequences are indicated as in Figures 1 and 2. Unless otherwise indicated, all names refer to human kinases. Whenever possible well-known genes or gene families are indicated as examples of the sequences included in each branch. Plant families were defined according to the PlantsP database (Tchieu et al. 2003).

**Figure 4.** Basic structures of the LRRK1 and LRRK2 proteins according to Pfam. The COR and LRRK2-specific repeats are not included in Pfam and have been positioned according to Bosgraaf and Van Haastert (2003) and this study.

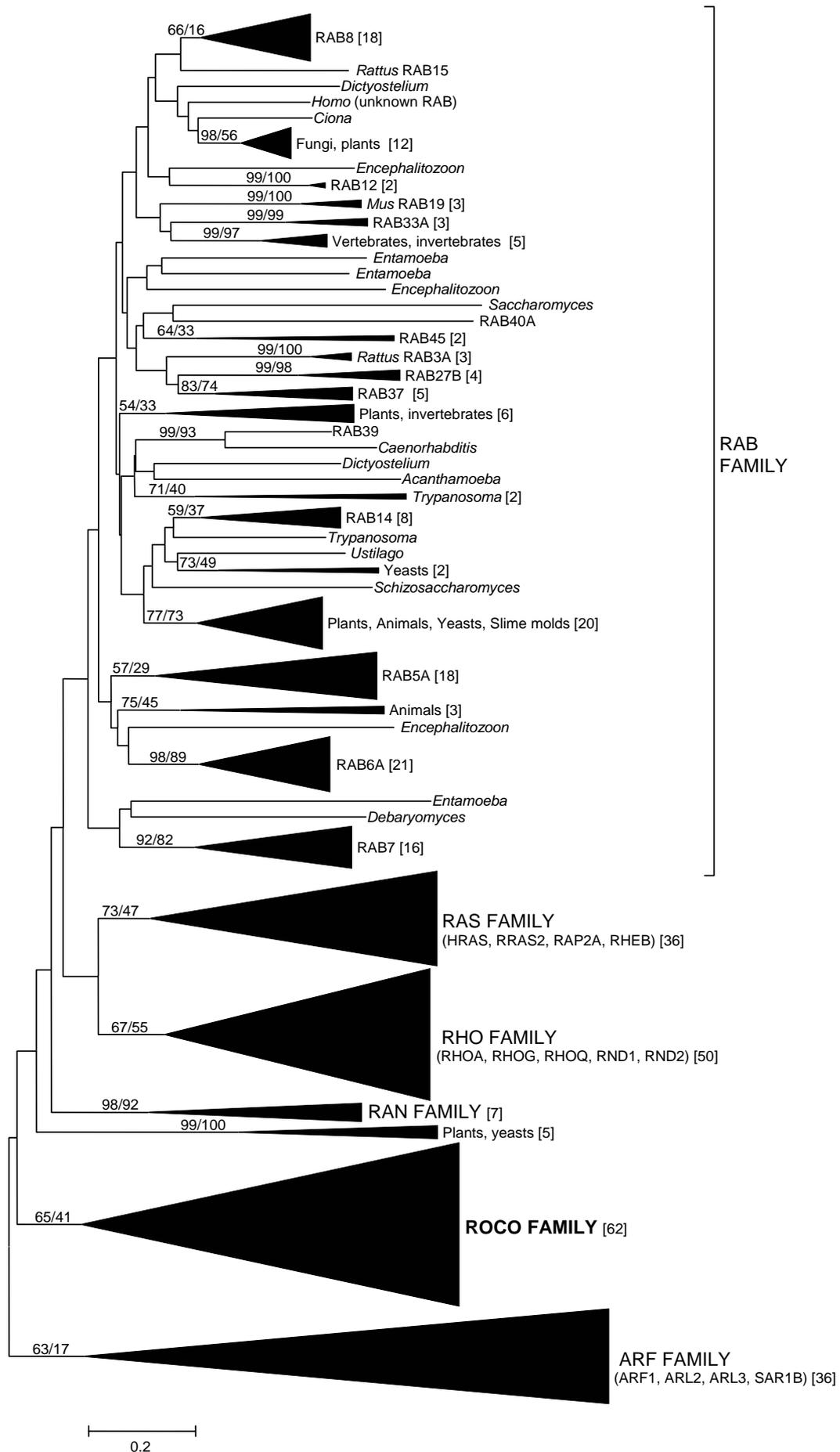
**Figure 5.** Sequences of the LRRK2-specific repeats in human LRRK2 (Top: 1-14) and in the LRRK2 orthologous protein found in the sea urchin *S. purpuratus* (Bottom: S1-S14). These repeats are 33-34 amino acids long.

**Figure 6.** Structures of the kinase domain of human B-raf (A), and models for human LRRK2 (B), human LRRK2 with the G2019S mutation (C) and human LRRK2 with the I2020T mutation (D). A strong modification of the local structure of the regions in which the mutations are introduced can be observed comparing panels C and D with panel B (asterisks).

**Figure 1. Marín**



**Figure 2. Marín**



**Figure 3. Marín**

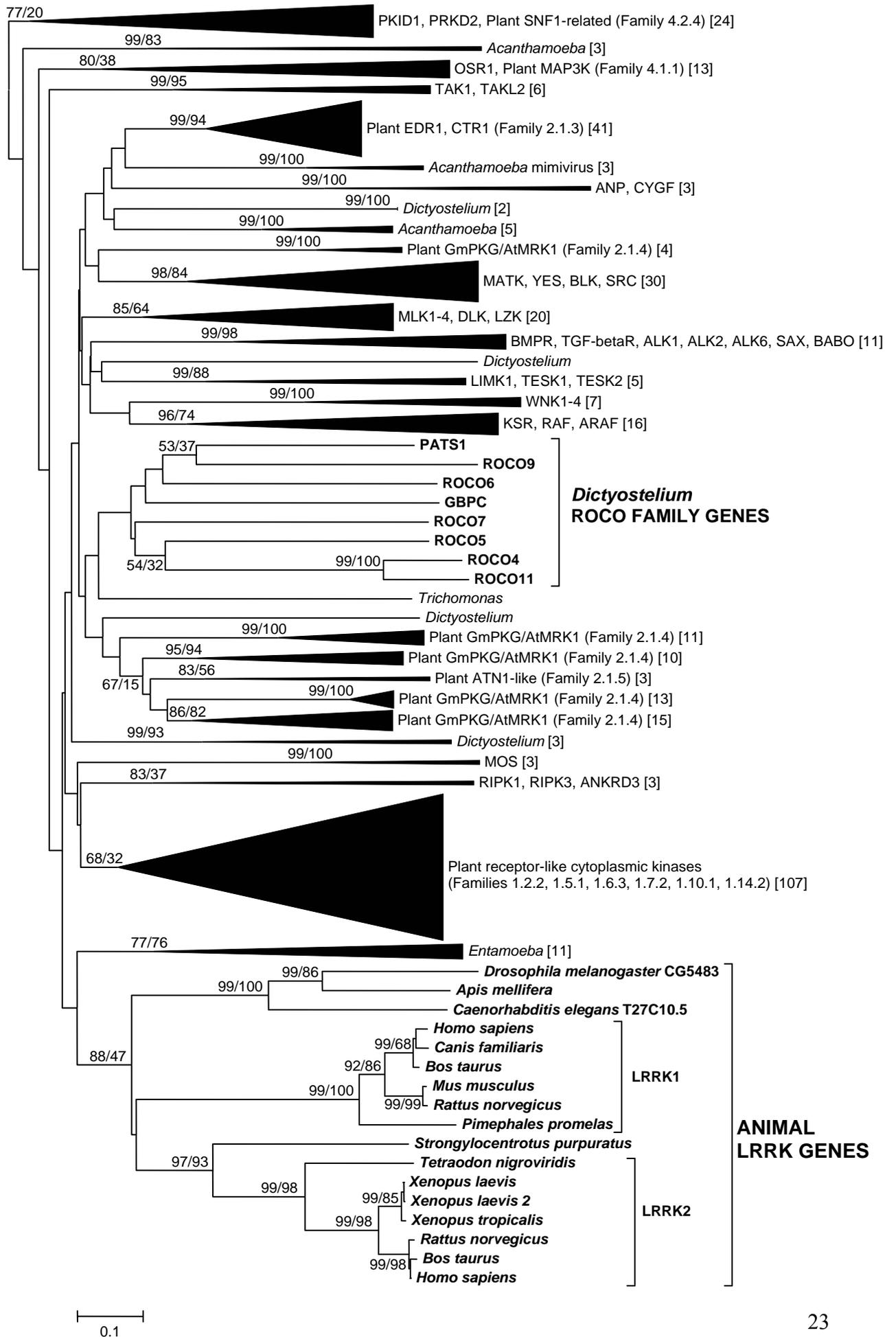
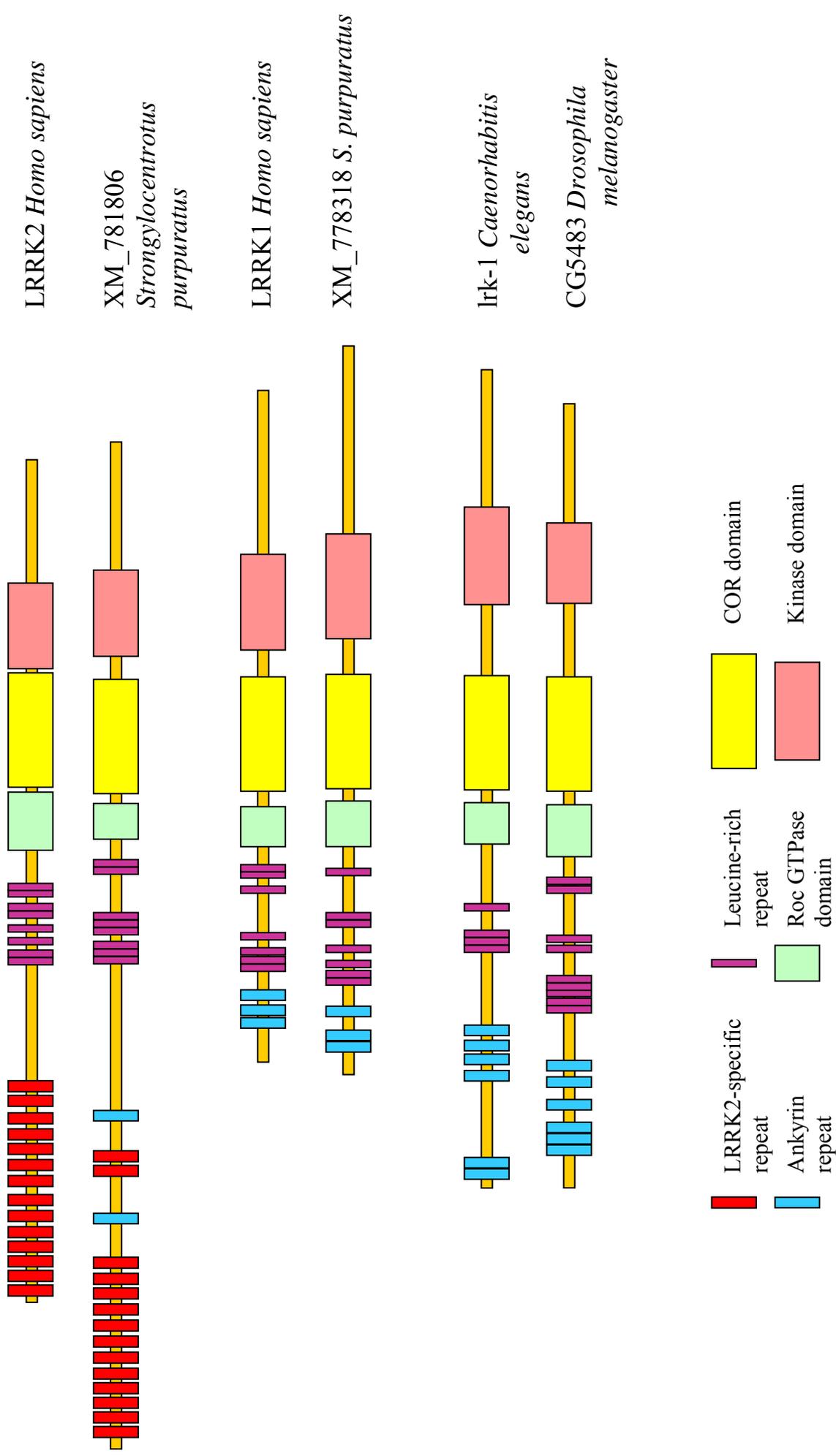


Figure 4. Marín





**Figure 6. Marín**

