

# The Parkinson Disease Gene *LRRK2*: Evolutionary and Structural Insights

Ignacio Marín

Departamento de Genética, Universidad de Valencia, Burjassot, Spain

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 these domains 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 that 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. Modeling based on known kinase structures suggests that mutations in *LRRK2* that cause familial PD may alter the local 3-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 recent years have witnessed the discovery of several genes the mutations of which strongly contribute to the generation of PD (see recent reviews by Abou-Sleiman et al. 2006; Farrer 2006). In October 2004, 2 studies demonstrated an involvement of mutations in the *leucine-rich repeat kinase 2* (*LRRK2*) gene in autosomal dominant familial PD (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 et al. 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 et al. 2006).

Before *LRRK2* was related to PD, a few researchers became interested in this gene because of its obvious relationship with several *Dictyostelium discoideum* genes involved in cytokinesis, cell polarity, and chemotaxis (Bosgraaf et al. 2002; Goldberg et al. 2002; Abe et al. 2003; Abysalh et al. 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 2 characteristic domains. The first, called Roc, is sim-

ilar 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 most common are typical leucine-rich repeats (LRRs), located N terminally with 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 et al. 2006).

The *LRRK2* gene is expressed in multiple tissues and in multiple brain regions in humans and rodents (Paisán-Ruiz et al. 2004; Zimprich 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 parallelism 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 effects 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 et al. 2006; Farrer 2006). These results, together with the fact that Lewy bodies and other proteinaceous inclusions are found in individuals affected by *LRRK2*

Key words: Parkinson disease, kinase, GTPase, Roco family.

E-mail: ignacio.marin@uv.es.

*Mol. Biol. Evol.* 23(12):2423–2433. 2006

doi:10.1093/molbev/msl114

Advance Access publication September 11, 2006

mutations (Wszolek et al. 2004; Zimprich 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 PD genes in order to provide novel hints about their cellular functions (Marín and Ferrús 2002; Marín et al. 2004; Lucas et al. 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 Roco and Rab families 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 some representative Roco kinase sequences plus a large number of sequences with high level of similarity to these kinase domains. Almost all of them belonged to the tyrosine kinase-like (TKL) class, to which Roco kinase 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.kinaset.org/pkr/>).

The protein sequences found were then aligned using ClustalX version 1.83 (Thompson et al. 1997), and preliminary 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, general analysis), 124 (Roc domain, specific analysis; see below), and 400 (kinase domain) sequences. From these databases, I generated the final multiple-protein alignments again using first ClustalX 1.83 and then GeneDoc 2.6 (Nicholas et al. 1997), to manually correct them. The sequences analyzed spanned the regions corresponding to amino acids 1337–1455 (Roc domain), 1519–1795 (COR domain), and 1886–2068 (kinase domain) in human *LRRK2*. Dendrograms were then obtained both by the NJ and the maximum-parsimony (MP) methods, using the routines available in MEGA 3.1 (Kumar et al. 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, and 4) heuristic search using the subtree pruning–regrafting algorithm. Support for the topologies obtained with those 2 methods was determined using the bootstrap routines also available in MEGA 3.1 and PAUP\*. In all, 1,000 replicates were performed for both NJ and MP bootstrap analyses. Dendrograms were depicted using the tree editor of MEGA 3.1.

### Structural Analyses

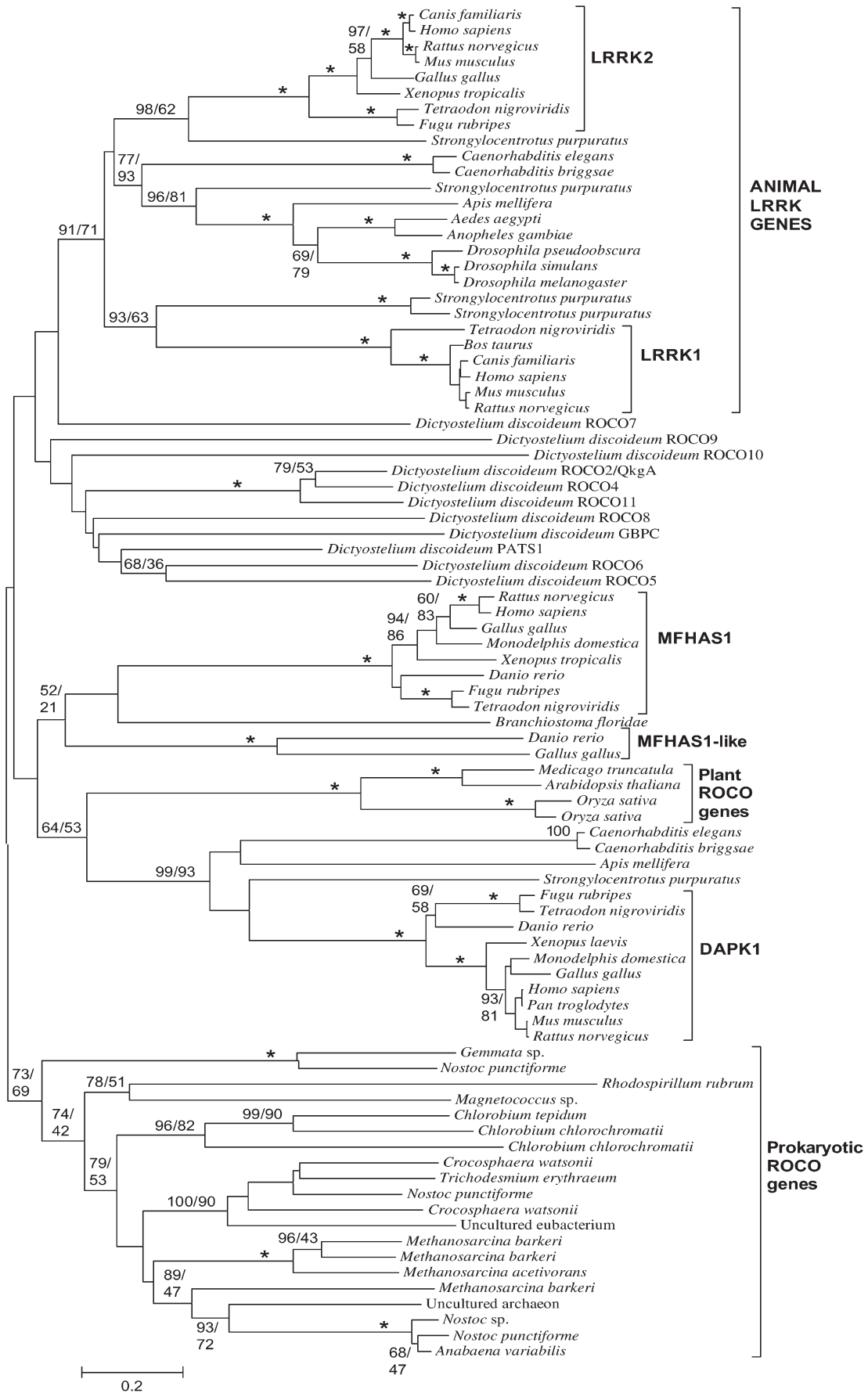
Domain searches were performed against the Pfam, SMART, and Conserved Domain databases, already cited. Motif searches with the human and sea urchin *LRRK2*-specific repeats were performed using PRATT 2.1 (Jonassen et al. 1995; <http://www.expasy.org/tools/pratt/>; minimum match: 25%) to generate sequence 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 3-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

FIG. 1.—Dendrogram 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 support the same branch and at least 1 of the 2 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.



NJ and MP 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 2 species, *Danio rerio* and *Gallus gallus* (see fig. 1). Finally, it is significant that the classification in 3 groups, based on domain architecture, proposed by Bosgraaf and Van Haastert (2003) is not supported by this phylogenetic analysis.

#### Singularity of the Roc GTPase Domain and Confirmation of the COR Results

Bosgraaf and Van Haastert (2003) performed a relatively limited phylogenetic analysis 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 a general 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 the previous findings of Bosgraaf and Van Haastert (2003), whereas 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.

Figure 3 shows a more specific analysis performed in order to determine whether the evolutionary history deduced in the previous section from the analyses of COR domains was also supported by similar analyses using

the Roc domains. As it can be easily seen by comparing figures 1 and 3, the main groups detected in the COR domain analyses are confirmed. However, because Roc domains are much shorter and therefore less informative than COR domains, in general, the bootstrap values to support those groups are smaller in figure 3 than those shown in figure 1. In the Roc domain–based tree, the *D. discoideum* Roco sequences appear closer to the sequences deduced from prokaryotic Roco genes than to animal *LRRK* sequences, although, again, there is no significant support for this alternative topology. There are some differences in the data sets used to generate figures 1 and 3 (e.g., there are 4 *S. purpuratus* sequences in fig. 1 and only 2 in fig. 3). They are due to the fact that, in many cases for which genome information is not complete, the sequences of particular domains can be reconstructed but the corresponding whole-length sequences of the genes are not available.

#### 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. These domains are, however, very different. Kinase domains in *LRRK* genes can be classified as belonging to the *TKL* group, whereas the *DAPK1* kinase domain can be included in the calcium/calmodulin-dependent kinase (*CAMK*) 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 *D. 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 cooptions of closely related kinase domains has never been tackled. Figure 4 shows the trees obtained for selected *LRRK* and *D. discoideum* kinase domains, plus their closest relatives (as appear in Blast searches), and a set of selected kinases of the *TKL* group. Notably, the animal and *Dictyostelium* sequences appear as 2 separated groups in distant positions in these trees. This result suggests that independent cooptions for related kinase domains have occurred in animal and *Dictyostelium* genes. The low bootstrap support for the inner branches of the tree precludes 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 with respect to the structures of LRRK proteins. Most authors depict human LRRK2 protein as having, going from the N-terminal to the C-terminal end, several LRRs (4 according to Bosgraaf and Van Haastert [2003], 13 according to Guo et al. [2006]); single Roc, COR, and kinase domains; and finally 1 (Mata et al. 2006), 2 (Bosgraaf and Van Haastert 2003), or even 7 (Guo et al. 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

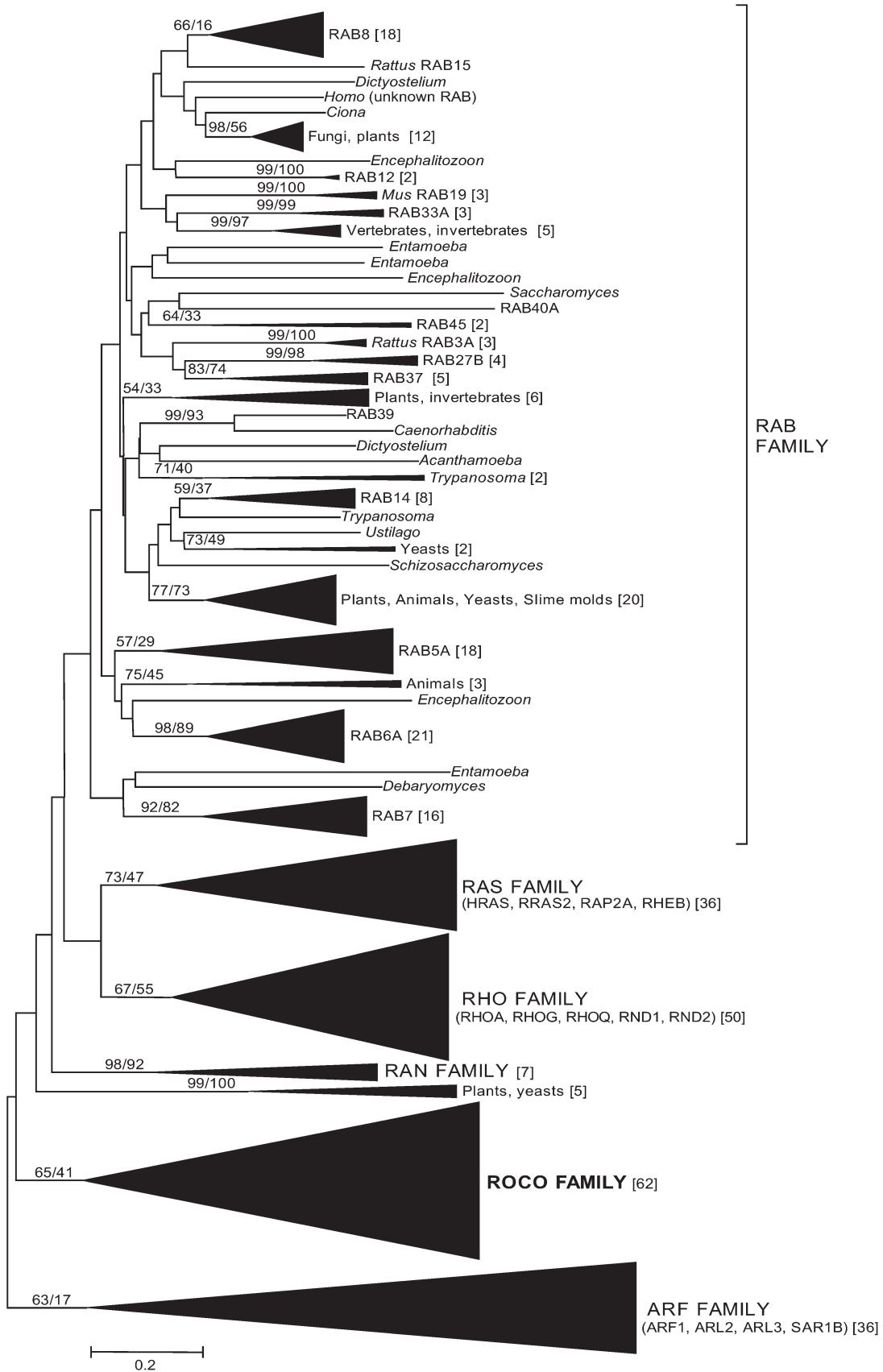


FIG. 2.—Tree showing the relationships of the main families of small GTPases of the Ras superfamily and the Roc GTPase domain found in Roco proteins. Bootstrap support shown as in figure 1. Numbers in brackets refer to the number of sequences in the branch.

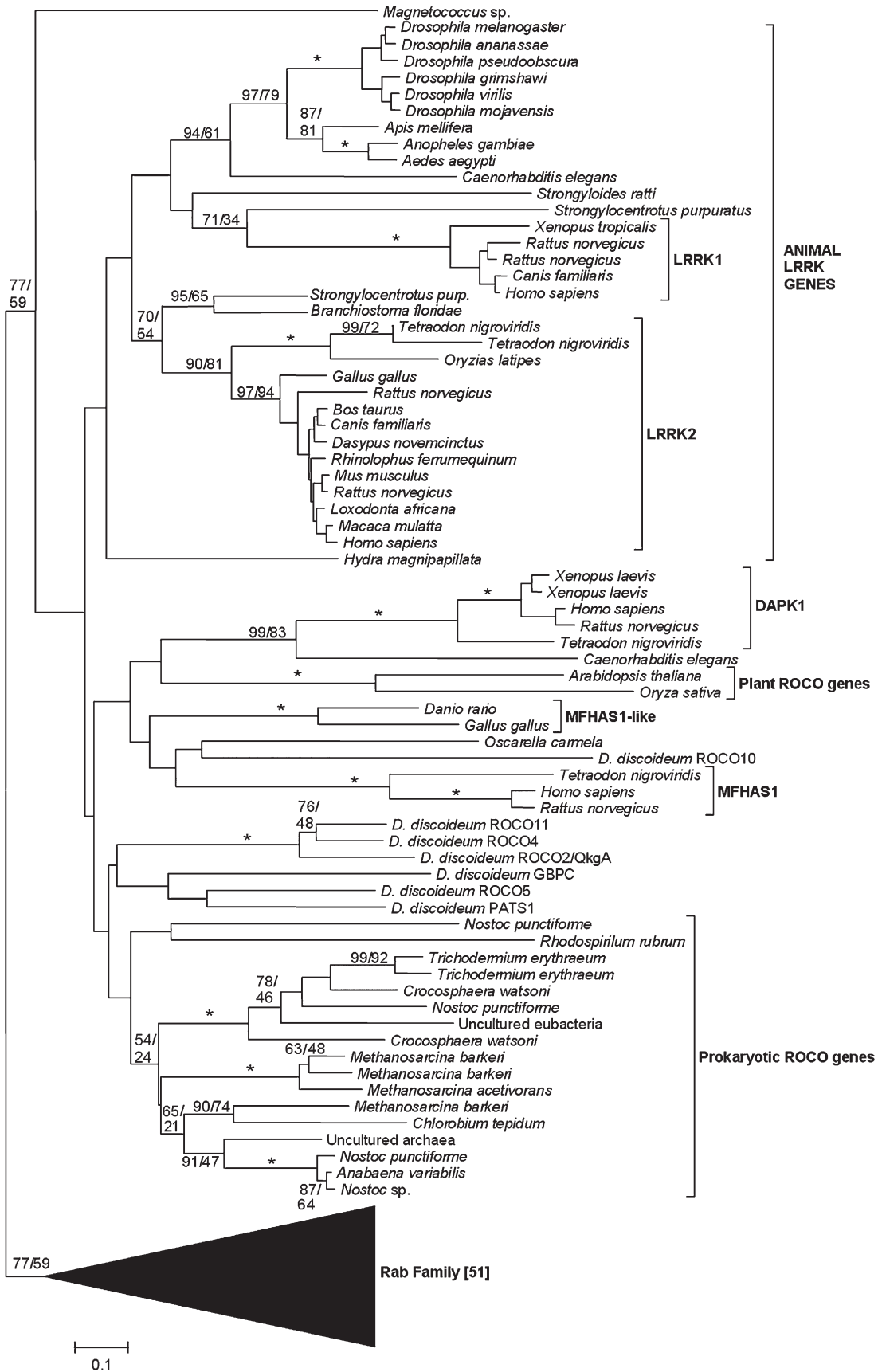


FIG. 3.—A precise analysis of the relationships among Roc domains using some Rab sequences as outgroups. Bootstrap support and number of sequences are detailed as in the previous figures. Again, a few significant values for short terminal branches have been omitted.

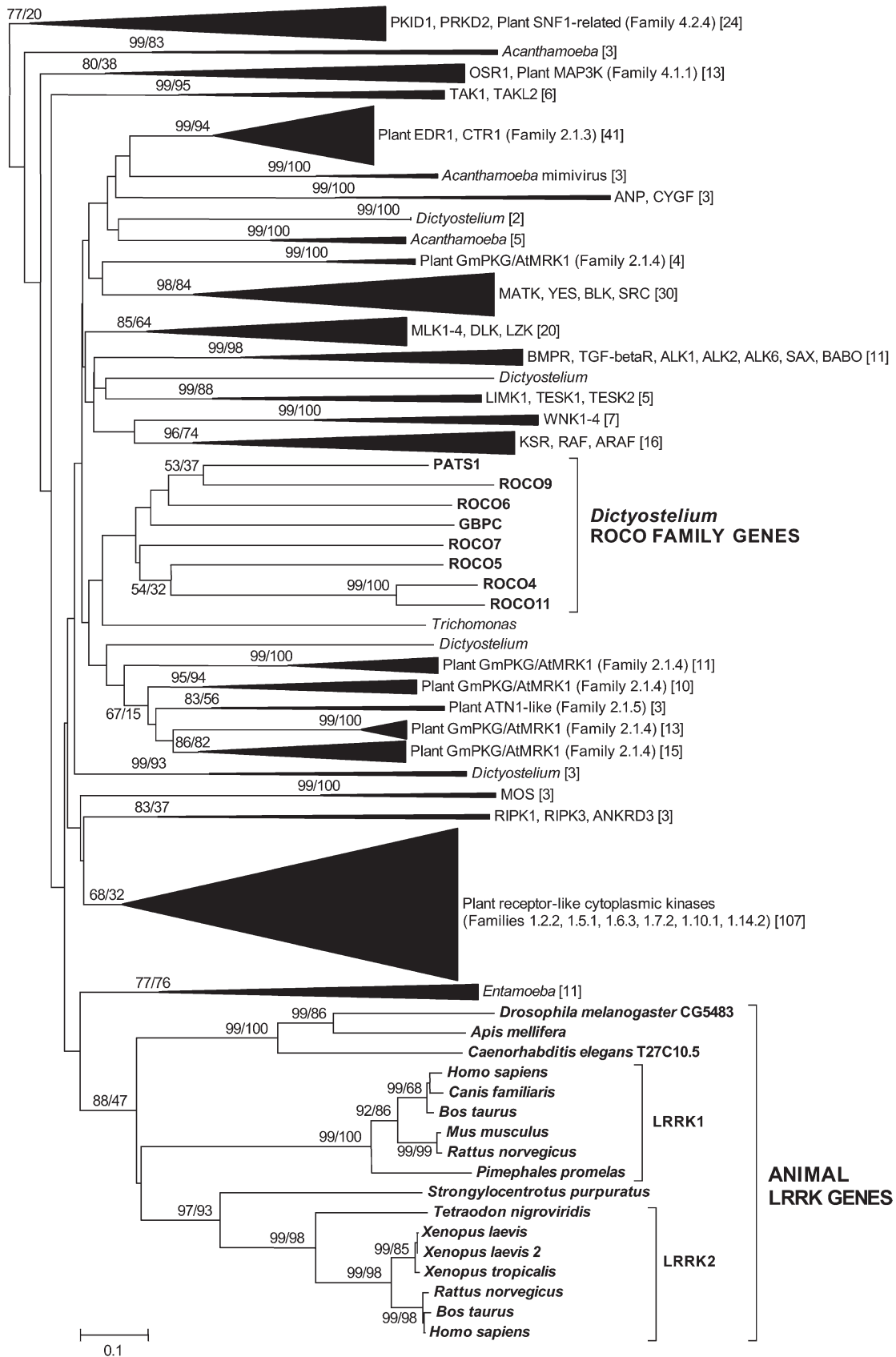


FIG. 4.—Dendrogram for kinase domains. Bootstrap values and number of sequences are indicated as in previous figures. Unless otherwise indicated, all names refer to human kinases. Whenever possible, well-known genes or gene families are indicated as examples of the sequences included in each branch. Plant families were defined according to the PlantsP database (Tchiew et al. 2003).

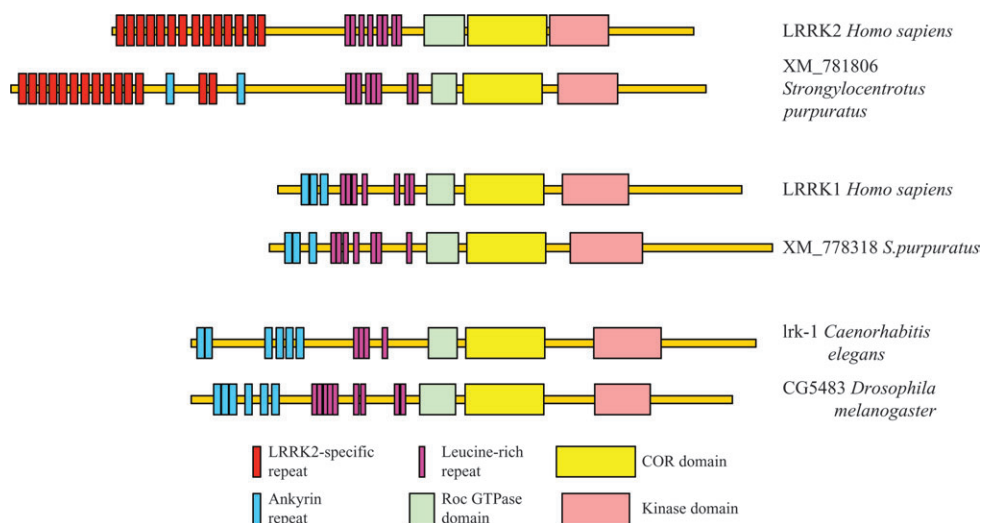


FIG. 5.—Basic structures of the LRRK1 and LRRK2 proteins according to Pfam. The COR and LRRK2-specific repeats are not included in Pfam and have been positioned according to Bosgraaf and Van Haastert (2003) and this study.

analyses using the Pfam, SMART, and Conserved domain (NCBI) databases support the existence in human LRRK2 proteins of several LRRs (7 according to SMART and 8 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 2231–2276 is predicted by SMART). No significant predictions for ankyrin repeats are ever found. 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 5 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 (fig. 5, 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 14 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 sequence of these repeats located along the first 660 (human) to 850 (sea urchin) amino acids of LRRK2 proteins are detailed in figure 6.

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 3-dimensional folding (e.g., Guo et al. 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 7, a model for the wild-type kinase domain of LRRK2 (fig. 7B) is shown, which closely resembles the structure of related kinases, such as human B-Raf (fig. 7A). 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 fig. 7C and D) 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 modeled (not shown).

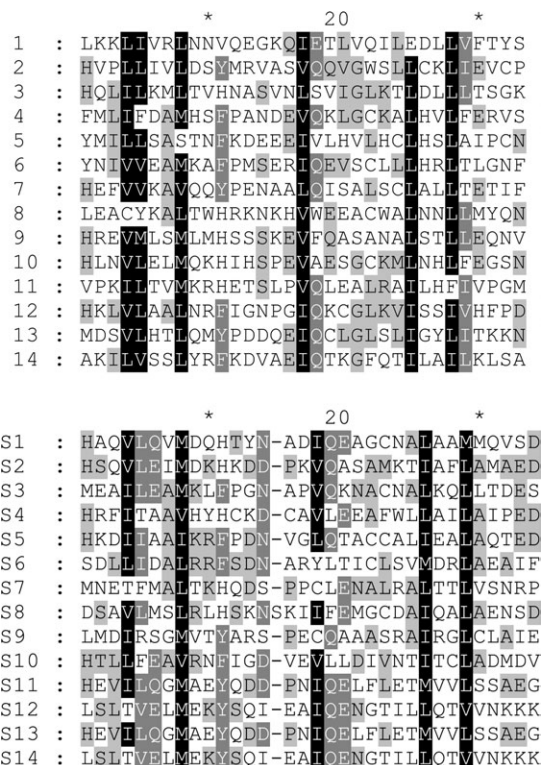


FIG. 6.—Sequences of the LRRK2-specific repeats in human LRRK2 (top: 1–14) and in the LRRK2 orthologous protein found in the sea urchin *Strongylocentrotus purpuratus* (bottom: S1–S14). These repeats are 33–34 amino acids long.

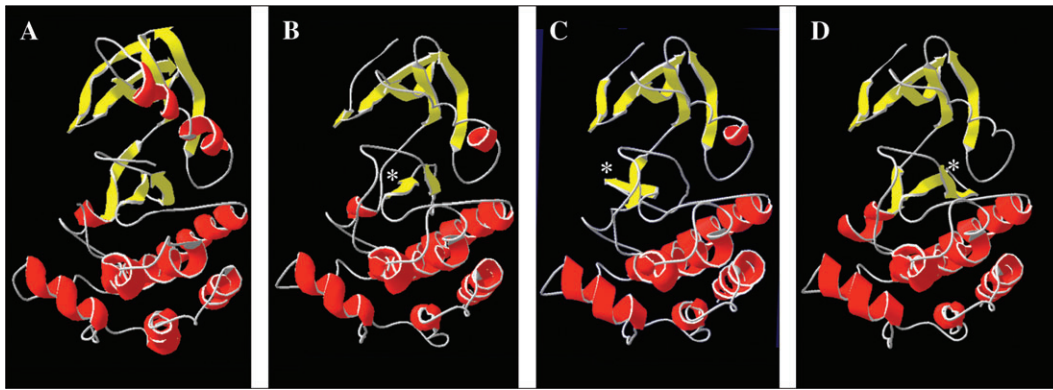


FIG. 7.—Structures of the kinase domain of human B-Raf (A) and models for human LRRK2 (B), human LRRK2 with the G2019S mutation (C), and human LRRK2 with the I2020T mutation (D). A strong modification of the local structure of the regions in which the mutations are introduced can be observed comparing panels (C) and (D) with panel (B) (asterisks).

## Discussion

This study is focused on describing the evolution and structural characteristics of Roco family genes, with emphasis on the PD gene *LRRK2*. To trace the general evolutionary history of this group of genes, analyses of the COR and Roc domain have been performed (see figs. 1 and 3). Roco family genes are present in prokaryotes, both eubacteria (cyanobacteria, proteobacteria, planctomycetes, and 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 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 independent cooptions of relatively similar kinase domains (fig. 4). Second, vertebrate-specific amplification of this family has occurred: protostomes have only 1–2 Roco genes, whereas up to 5 can be found in vertebrates. COR and Roc domain analyses have shown that *LRRK2* emerged by gene duplication quite recently, after the protostome–deuterostome split (figs. 1 and 3). Moreover, protostome LRRK proteins are structurally much more similar to deuterostome LRRK1 proteins than to deuterostome LRRK2 proteins (fig. 5). 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 et al. 2006). Data from the Roc (fig. 2) and kinase domains (fig. 4) show that they are quite different from any other GTPases or kinases found in the databases, warning against simplistic views of LRRK2 proteins as fusions 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 figs. 2 and 4). The complex structures of Roco proteins, and most especially of those encoded by *LRRK2* genes (fig. 5), 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 2 mutations in the kinase domain most likely generate significant local changes of the 3-dimensional structure of that domain but without affecting their overall folding (fig. 7). These results are compatible with the mutant kinase domains being active, as shown in recent experiments (West et al. 2005; Gloeckner et al. 2006).

## Supplementary Material

The alignments and trees shown in figures 1–4 can be obtained as a compressed file entitled “Marín—alignments and trees.rar” at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgment

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

## Literature Cited

- Abe T, Langenick J, Williams JG. 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.
- Abou-Sleiman PM, Muqit MMK, Wood NW. 2006. Expanding insights of mitochondrial dysfunction in Parkinson's disease. *Nat Rev Neurosci.* 7:207–219.
- Abysalh JC, Kuchnicki LL, Larochelle DA. 2003. The identification of *pats1*, a novel gene locus required for cytokinesis in *Dictyostelium discoideum*. *Mol Biol Cell.* 14:14–25.
- Berg D, Schweitzer K, Leitner P, Zimprich A, Lichtner P, Belcredi P, Brussel T, Schulte C, Maass S, Nagele T. 2005. Type and frequency of mutations in the LRRK2 gene in familial and sporadic Parkinson's disease. *Brain.* 128:3000–3011.
- Bosgraaf L, Russcher H, Smith LJ, Wessels D, Soll DR, Van Haastert PJ. 2002. A novel cGMP signalling pathway mediating myosin phosphorylation and chemotaxis in *Dictyostelium*. *EMBO J.* 21:4560–4570.
- Bosgraaf L, Van Haastert PJM. 2003. Roc, a Ras/GTPase domain in complex proteins. *Biochem Biophys Acta.* 1643:5–10.
- Deiss LP, Feinstein E, Berissi H, Cohen O, Kimchi A. 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 MJ. 2006. Genetics of Parkinson disease: paradigm shifts and future prospects. *Nat Rev Genet.* 7:306–318.
- Galter D, Westerlund M, Carmine A, Lindqvist E, Sydow O, Olson L. 2006. LRRK2 expression linked to dopamine-innervated areas. *Ann Neurol.* 59:714–719.
- Gattiker A, Gasteiger E, Bairoch A. 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl Bioinformatics.* 1:107–108.
- Giasson BI, Covy JP, Bonini NM, Hurtig HI, Farrer MJ, Trojanowski JQ, Van Deerlin VM. 2006. Biochemical and pathological characterization of Lrrk2. *Ann Neurol.* 59:315–322.
- Gilks WP, Abou-Sleiman PM, Gandhi S, et al. (15 co-authors). 2005. A common LRRK2 mutation in idiopathic Parkinson's disease. *Lancet.* 365:415–416.
- Gloeckner CJ, Kinkl N, Schumacher A, Braun RJ, O'Neill E, Meitinger T, Kolch W, Prokisch H, Ueffing M. 2006. The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. *Hum Mol Genet.* 15:223–232.
- Goldberg JM, Bosgraaf L, Van Haastert PJ, Smith JL. 2002. Identification of four candidate cGMP targets in *Dictyostelium*. *Proc Natl Acad Sci USA.* 99:6749–6754.
- Goldberg JM, Manning G, Liu A, Fey P, Pilcher KE, Xu Y, Smith JL. 2006. The *Dictyostelium* kinome—analysis of the protein kinases from a simple model organism. *PLoS Genet.* 2:e38.
- Greggio E, Jain S, Kingsbury A, et al. (18 co-authors). 2006. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol Dis.* 23:329–341.
- Guex N, Peitsch MC. 1997. Swiss-Model and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 18:2714–2723.
- Guo L, Wang W, Chen SG. 2006. Leucine-rich repeat kinase 2: relevance to Parkinson's disease. *Int J Biochem Cell Biol.* 38:1469–1475.
- Jonassen I, Collins JF, Higgins DG. 1995. Finding flexible patterns in unaligned protein sequences. *Protein Sci.* 4:1587–1595.
- Korr D, Toschi L, Donner P, Pohlentz HD, Kreft B, Weiss B. 2006. LRRK1 protein kinase activity is stimulated upon binding of GTP to its Roc domain. *Cell Signalling.* 18:910–920.
- Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinformatics.* 5:150–163.
- Lucas JI, Arnau V, Marín I. 2006. Comparative genomics and protein domain graph analyses link ubiquitination and RNA metabolism. *J Mol Biol.* 357:9–17.
- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. 2002. The protein kinase complement of the human genome. *Science.* 298:1912–1934.
- Marco A, Cuesta A, Pedrola L, Palau F, Marín I. 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.
- Marín I, Ferrús A. 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, Lucas JI, Gradilla AC, Ferrús A. 2004. Parkin and relatives: the RBR family of ubiquitin ligases. *Physiol Genomics.* 17:253–263.
- Mata IF, Wedemeyer WJ, Farrer MJ, Taylor JP, Gallo KA. 2006. LRRK2 in Parkinson's disease: protein domains and functional insights. *Trends Neurosci.* 29:286–293.
- Melrose H, Lincoln S, Tyndall G, Dickson D, Farrer M. 2006. Anatomical localization of leucine-rich repeat kinase 2 in mouse brain. *Neuroscience.* 139:791–794.
- Nicholas KB, Nicholas HB Jr, Deerfield DW. 1997. GeneDoc: analysis and visualization of genetic variation. *EMBNEW News.* 4:14.
- Paisán-Ruiz C, Jain S, Evans EW, et al. (21 co-authors). 2004. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron.* 44:595–600.
- Peitsch MC. 1996. ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem Soc Trans.* 24:274–279.
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 4:406–425.
- Sakabe T, Shinomiya T, Mori T, Ariyama Y, Fukuda Y, Fujiwara T, Nakamura Y, Inazawa J. 1999. Identification of a novel gene, *MASLI*, within an amplicon at 8p23.1 detected in malignant fibrous histiocytomas by comparative genomic hybridization. *Cancer Res.* 59:511–515.
- Simon-Sanchez J, Herranz-Perez V, Olucha-Bordonau F, Perez-Tur J. 2006. LRRK2 is expressed in areas affected by Parkinson's disease in the adult mouse brain. *Eur J Neurosci.* 23:659–666.
- Skipper L, Li Y, Bonnard C, et al. (11 co-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 WW, Pei Z, Jiang H, Moore DJ, Liang Y, West AB, Dawson VL, Dawson TM, Ross CA. 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 DL. 2003. PAUP\*: phylogenetic analysis using parsimony (\*and other methods). Version 4. Sunderland (MA): Sinauer Associates.
- Tan EK, Skipper L, Chua E, Wong MC, Pavanni R, Bonnard C, Kolatkar P, Liu JJ. 2006. Analysis of 14 LRRK2 mutations in Parkinson's plus syndromes and late-onset Parkinson's disease. *Mov Disord.* 21:997–1001.
- Taylor JP, Mata IF, Farrer MJ. 2006. LRRK2: a common pathway for parkinsonism, pathogenesis and prevention? *Trends Mol Med.* 12:76–82.

- Tchieu JH, Fana F, Fink JL, et al. (12 co-authors). 2003. The PlantsP and PlantsT functional genomics databases. *Nucleic Acids Res.* 31:342–344.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 24:4876–4882.
- West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, Dawson VL, Dawson TM. 2005. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc Natl Acad Sci USA.* 102:16842–16847.
- Wszolek ZK, Pfeiffer RF, Tsuboi Y, et al. (13 co-authors). 2004. Autosomal dominant parkinsonism associated with variable synuclein and tau pathology. *Neurology.* 62:1619–1622.
- Zimprich A, Biskup S, Leitner P, et al. (22 co-authors). 2004. Mutations in *LRRK2* cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron.* 44:601–607.

Stephanie Monks, Associate Editor

Accepted September 7, 2006