Evolution of the Rho family of ras-like GTPases in eukaryotes.

Anthony Boureux, Emmanuel Vignal, Sandrine Faure, Philippe Fort

Evolution of the Rho family of ras-like GTPases in eukaryotes

Boureux Anthony , Vignal Emmanuel , Faure Sandrine , Fort Philippe *

CRBM, Centre de recherches de biochimie macromolé*culaire CNRS : FRE2593, IFR122, 1919 Route de Mende 34293 MONTPELLIER CEDEX 5,FR

* Correspondence should be adressed to: Philippe Fort <philippe.fort@crbm.cnrs.fr>

Abstract

GTPases of the Rho family are molecular switches that play important roles in converting and amplifying external signals into cellular effects. Originally demonstrated to control the dynamics of the F-actin cytoskeleton, Rho GTPases have been implicated in many basic cellular processes that influence cell proliferation, differentiation, motility, adhesion, survival or secretion. To elucidate the evolutionary history of the Rho family, we have analyzed over twenty species covering major eukaryotic clades from unicellular organisms to mammals, including platypus and opossum, and have reconstructed the ontogeny and the chronology of emergence of the different subfamilies. Our data establish that the 20 mammalian Rho members are structured into eight subfamilies, among which Rac is the founder of the whole family. Rho, Cdc42, RhoUV and RhoBTB subfamilies appeared before Coelomates, and RhoJQ, RhoDF and Rnd emerged in Chordates. In Vertebrates, gene duplications and retrotranspositions increased the size of each chordate Rho subfamily, while RhoH, the last subfamily, arose probably by horizontal gene transfer. Rac1b, a Rac1 isoform generated by alternative splicing, emerged in amniotes, and RhoD, only in therians. Analysis of Rho mRNA expression patterns in mouse tissues shows that recent subfamilies have tissue-specific specific and low level expression, which supports their implication only in narrow time windows or in differentiated metabolic functions. These findings give a comprehensive view of the evolutionary canvas of the Rho family and provide guides for future structure and evolution studies of other components of Rho signaling pathways, in particular regulators of the RhoGEF family.

MESH Keywords Amino Acid Sequence ; Animals ; Evolution ; Molecular ; Fungi ; genetics ; Gene Duplication ; Humans ; Invertebrates ; genetics ; Molecular Sequence Data ; Phylogeny ; Plants ; genetics ; Pseudogenes ; Sequence Alignment ; Vertebrates ; genetics ; rho GTP-Binding Proteins ; genetics

Introduction

Development of multicellular organisms requires an extraordinary “sensing” ability of cells to detect and respond adequately to cues expressed by other cells (adhesion molecules, extracellular matrix, cytokines, morphogens, growth factors or hormons). Inter-cellular signaling was extensively studied in dynamic situations such as embryonic development and the use of simple genetic models has allowed the identification of pathways highly conserved in most eukaryotes. Cell signaling is initiated by the binding of ligands to their receptors at the cell surface, and then converted into specific responses, which mostly affect gene transcription, cell shape, adhesion, motility, and endo/exocytosis. Since the identification of the first member Ha-Ras as a viral 21 kDa protein responsible for tumor formation (Andersen et al. 1981), Ras and related members have been found in all studied eukaryotic organisms and are probably the most conserved proteins amongst the cellular components involved in cell signaling. Ras-like proteins usually are low molecular weight proteins that display a conserved structural backbone of five G-boxes involved in GTP binding and GTPase activity (Bourne, Sanders, and McCormick 1991). Most Ras-like GTPases act as signaling gates, which are switched on when bound to GTP and off when bound to GDP. The switch is positively controlled by Guanine nucleotide Exchange Factors (GEF), which catalyze the replacement of GDP by GTP and negatively by GTPase Activating Proteins (GAP), which accelerate the intrinsic GTPase activity thereby favoring the GDP bound form. When bound to GTP, the GTPase gets an active conformation and interacts with effectors that mediate downstream cellular effects. Ras-like proteins constitute a super-family of over 150 members in mammals, subdivided into five main families: Ras, Rho, Rab, Arf and Ran, which control each particular aspects of cell metabolism, such as cell proliferation for Ras (Hancock and Parton 2005; Wennerberg, Rossman, and Der 2005), cell morphology for Rho (Wennerberg and Der 2004), vesicle trafficking for Rab and Arf (Donaldson and Honda 2005; Bucci and Chiariello 2006) and nuclear trafficking for Ran (Pemberton and Paschal 2005).

Rho family members (Madaule and Axel 1985) are defined by the presence of a Rho-specific specific insert located between the G4 and G5 boxes and involved in the binding to effectors and regulators (Freeman, Abo, and Lambeth 1996). Like other Ras-like, Rho proteins are present from lower eukaryotes such as the slime mold and yeast (Tanaka and Takai 1998; Rivero et al. 2001) up to mammals (Wennerberg and Der 2004). First described as promoting reorganization of the F-actin cytoskeleton (Hall 1998), Rho proteins have been shown to also participate to many pathways that affect cell proliferation, apoptosis, adhesion, motility and differentiation, gene expression and vesicular trafficking (Ridley 2001). In mammals, the Rho family contains about 20 members structured into subfamilies (Wherlock and Mellor 2002), but most functional data pertained to Rac, Rho and Cdc42 only. The physiological functions and ontogeny of most members thus remain poorly understood.

The aim of the present study was to compare Rho families amongst eukaryotic clades to get an insight into the evolutionary history of each subfamily. Such analysis had never been done because of the low number of eukaryotic genome projects completed so far, and we took here opportunity of genomic data from taxons that cover most eukaryotic clades over 1.5 billion years. We have examined the complete Rho families in 26 eukaryotic genomes, including the most recent ones (hemichordates, echinoderms and prototherians), reconstructed the ontogeny of each Rho subfamily and specified the timing of their emergence. While supporting the pivotal roles of Rac, Rho and Cdc42, our data give a different picture on the evolution of other members and their potential physiological roles.

Material and Methods

Database searches

We searched genomic and/or EST databases for Rho GTPases using TBLASTN or BLASTP (v2.2.13) algorithms (Altschul et al. 1997). Searches were done either on remote servers (Ensembl, PlasmoDB, TIGR, Sanger Institute, JGI, CiliateDB and NCBI) or on a standalone PowerPC G5 computer (Apple). Downloaded genomic sequences were assembled using ABI Prism Auto-Assembler (v2.1, Perkin Elmer). Hits from searches in annotated databases (Ensembl) were checked for appropriate translation and corrected in most cases. Protein sequences and gene features are shown Table S1 (supplemental data). We searched in the following organisms: Fungi: Aspergillus fumigatus, Cryptococcus neoformans, Yarrowia lipolytica, Ustilago maydis, Entamoebidae: Entamoeba hystolytica, Alveolates: Plasmodium falciparum, Tetrahymena thermophila, Stramenopiles: Phytophthora ramorum, Thalassiosira pseudonana, Porifera: Reniera sp. JGI-2005, Cnidarians: Hydra magnipapillata, Acoelomates: Schisostoma japonicum and Schistosoma manson, Hemichordates: Saccoglossus kowaleski, Echinoderms: Strongylocentrotus purpuratus, Urochordates: Oikopleura dioica, Molgula tectiformis, Cephalochordates: Branchiostoma floridae, Vertebrates: Danio rerio, Takifugu rubripes, Tetraodon nigroviridis, Xenopus tropicalis, Xenopus laevis, Gallus gallus, Ornithorhynchus anatinus, Monodelphis domestica, Loxodonta africana, Bos taurus, Canis familiaris, Mus domesticus, Rattus norvegicus and Homo sapiens. Classification and genome projects web URLs are summarized Table S2 (supplemental data).

Protein alignment and phylogenetic analysis

Sequences restricted to the core Rho domain (i.e. aminoacids 5-173 in Rac1) were aligned using ClustalX (Jeanmougin et al. 1998) with BLOSUM30 alignment matrix. Rac1 secondary structure was used to set local gap penalties to keep G1 to G5 GTP-binding boxes aligned. Unrooted trees were derived from optimized alignments using bootstrap neighbor joining (Clustal X 1.83, seed=111, N=1000) or maximum likelihood (ProML 3.6.3, J. Felsenstein, University of Washington) (Saitou and Nei 1987; Felsenstein 1996). Trees were displayed using TreeView (Page 1996) and edited in Adobe Illustrator CS. Selective constraints on RhoD and RhoF protein sequences were adressed by computation of synonymous (Ks) and non-synonymous (Ka) mutation rates using the DnaSP package (Rozas et al. 2003).

SAGE analysis

We collected more than 3.8 million experimental tags (with 1143637 unique tags) from 244 publicly available mouse SAGE libraries retrieved from the SAGE Genie repository (Boon et al. 2002). All SAGE and Tag-to-gene mapping informations from SAGE Genie were parsed and inserted into a relational database. Regular SAGE Rho gene tags were identified using the best_tag information provided by SAGE Genie and are listed in Table S3. For all libraries, tag informations (including tag per million) for each Rho gene were extracted from the database (available on request on tabular file format). Only tags found at least twice in libararies were considered. The spreadsheet OpenOffice Calc program was used for the analysis.

Results

Definition of Rho family sub-classes and members

Since the identification of RhoA in 1985, about twenty related Rho members have been identified in the human genome, the first vertebrate genome to be completed (Venter et al. 2001). The understanding of the Rho family structure remained nonetheless blurred, mainly because of lack of accurate phylogenetic analysis and nomenclature inconsistency. Using CLUSTALX neighbor-joining and ProML maximum likelihood methods, we reexamined the Rho phylogeny and confirmed the presence of eight subgroups distributed into four unambiguous clusters, supported by bootstrap values above 70% (Figure 1): The cluster I which contains the Rho (A-C), Rnd (1-3) and RhoD/RhoF subgroups, the cluster II, made of Rac/RhoG, Cdc42/RhoJ/RhoQ and RhoU/RhoV subgroups, the cluster III (RhoH) and cluster IV (RhoBTB1-2). Our analysis rejected the branching of MIRO and RhoBTB3 proteins as genuine Rho family members. MIRO proteins indeed confidently branched out before the Rho stem and should be considered as an autonomous Ras-like subfamily. The position of MIRO outside the Rho family is supported by the absence of Rho insert and by the equal similarity to Rho and Rab proteins (<45%, $p = 10^{-12}$). RhoBTB3 showed an equally low similarity score to Rho and Ras proteins (<45%, $p=10^{-4}$) but over a region of 100 amino acids only and should not thus be included in the family, even though the COOH moiety is related to the bona fide Rho RhoBTB1 and 2. We thus restricted the following analysis to the genuine 20 human Rho GTPase homologues.

Rho members in eukaryotes up to Bilaterian

Rho GTPases are absent in eubacteria and archae and are specific of eukaryotes. Rho families were identified previously in several eukaryotic kingdoms: 5 Rho and Cdc42 in S. cerevisiae (fungi) (Tanaka and Takai 1998), 13 Rop (related to Rac) in A. thaliana (plants) (Valster, Hepler, and Chernoff 2000), 15 Rac and RhoBTB in D. discoideum (mycetozoans) (Rivero et al. 2001). However, the D. discoideum RhoBTB (Rivero et al. 2001) is related to Rac and not to the metazoan RhoBTB. We searched for Rho genes in available sequence data of unicellular eukaryotes and found the presence of Rho and Cdc42 genes in most fungi [http://www.broad.mit.edu/annotation/fgi/], as well as Rac-like sequences in entamoeba (Entamoeba histolytica), in alveolates (the ciliate Tetrahymena thermophila, GenBank CX586341 and CH445588) and in stramenopiles (Phytophthora ramorum, orf 54454). Whereas absent in S. cerevisiae, we found Rac genes in several other fungi, such as Yarrowia lipolytica (XP_504400.1), Ustilago maydis (AACP01000023.1), Aspergillus fumigatus (AAHF01000002) or Cryptococcus neoformans (NC_006682). In contrast, we found no Rho member in the alveolate Plasmodium falciparum or in the stramenopile diatom Thalassiosira pseudonana. Rho evolution in these species is illustrated in Figure 2 and shows that Rop and Cdc42 clusters are both embedded into the Rac subgroup. This supports a scenario in which Rac genes have spread during eukaryotic crown radiation (i.e. more than 1.5 billion years ago, (Hedges et al. 2004)) and probably are the founders of the Cdc42 and Rop subfamilies, which constitute clearly identified clusters. The situation is less clear for the Rho subgroup, which forms a more diffuse cluster branched close to the root (delineated by the RhoBTB sequences). Either Rho diverged from Rac before Cdc42 in the clade leading to fungi and metazoans or it emerged earlier and was lost in the other clades.

We next examined the Rho family in three eumetazoan clades (Table 1): six members in the demosponge Reniera sp. JGI-2005 (Rho and Rac (1-5)) and in the hydrozoan Hydra magnipapillata (Cdc42, Rac, Rho (1-3) and RhoBTB, http://cnidbase.bu.edu/) and eight members in the acoelomates Schistosoma mansoni and japonicum (Cdc42, Rac (1-2) and Rho (1-5)). The Rho repertoire thus remained very similar in number and complexity from unicellular eukaryotes to primitive metazoan. Rho families are mainly made of duplicated Rho or Rac genes, which

indicates that the emergence of cell to cell interactions was not associated with new Rho members. These data also enlighten the high dynamics of the family in terms of expansion (e.g. Rac in mycetozoans, entamoebidae and plants, Rho in yeast, sponge or schistosoma) or loss (e.g. Rac in yeast and in plasmodium, Cdc42 in sponges and probably RhoBTB in sponges and schistosoma).

Emergence of Mtl and RhoUV subfamilies in Coelomates

We next addressed the evolution of the Rho complexity in coelomates by analyzing the ecdysozoan D. melanogaster and C. elegans (8 and 7 members, respectively, ENSF00000000175 and ENSF00000002177 ensembl protein families) and two primitive deuterostomians (cDNAs from the hemichordate acorn worm Saccoglossus kowalevskii and genome of the echinoderm sea urchin Strongylocentrotus purpuratus), from which we identified 7 S.kow and 11 S.pur. Rho sequences (Table 1). The clustering analysis of acorn worm (Sk), sea urchin (Sp), fly (Dm) and nematode (Ce) Rho sequences with those of hydra (Hm) and human (Hs) is shown in Figure 3A. The analysis produced six significant clusters: i) RhoA, Rac, and Cdc42, found in all examined species, in keeping with their presence in lower eukaryotes, and RhoBTB, noticeably absent in C. elegans and lower eukaryotes except hydra (Table 1). We did not found in any species a Cdc42 splice variant, as it is the case in mammals (Marks and Kwiatkowski 1996) ii) Mtl, a Rac/Cdc42 sibling cluster absent in hydra, schistosoma and present in ecdysozoans, hemichordates and echinoderms and lost in human. iii) RhoU, found in all deuterostomian species but also in fly (CG12102) and nematode (F22E12.2), a feature unnoticed so far (Wherlock and Mellor 2002). The clustering is supported by the presence of eight synapomorphic positions, which discriminate RhoU from the Rac and Cdc42 members (Figure 3B). These positions were also found in the mosquito and honey bee orthologues (ENSANGP00000028959 and ENSAPMP00000018001, not shown). The fruitfly RhoU (DmCG12102) exhibits a putative unconventional " Cxx" carboxy-terminal motif, responsible for membrane localization in human RhoU and RhoV (Berzat et al. 2005). DmRhoU is thus probably fully functional, but this remains to be experimentally settled. The nematode CeRhoU (F22E12.2 locus) showed numerous apomorphic states (Figure 3B), in particular a G12A substitution (Ras numbering) shown to be critical for Ras activity (Seeburg et al. 1984). In addition, CeRhoU lacks the amino-terminal terminal extension, the Rho-specific insert and the carboxy-terminal CAAX-box, which suggests that CeRhoU may now be inactive. This also suggests that either CeRhoU was submitted to particular evolutionary events which led to the loss of Rho-specific hallmarks or more likely, its clustering to the RhoU subfamily resulted from homoplasy.

Emergence of RhoJQ, Rnd, RhoDF and Cdc42b in Chordates

We previously reported the identification of the Rho family in the sea squirts Ciona intestinalis (C.int) and Ciona savignyi (C. sav), in which RhoJQ and RhoDF members were found, as well as two alternatively spliced Cdc42 isoforms (Philips et al. 2003). To extend Rho analysis in chordates, we examined Rho members in the sub-phylum cephalochordates (the lancelet Branchiostoma floridae, B. flo) and in other urochordates (the stolidobranch Molgula tectiformis, M. tec. and the appendicularian Oikopleura dioica, O. dio.). As shown in Table 2, the lancelet Rho family contained RhoJQ and Rnd related members in addition to the classical RhoABC, Rac, Cdc42, RhoBTB and RhoU, but contained no RhoDF member nor a Cdc42 splice variant. The picture differed in C. int, in which RhoJQ, RhoDF, a Cdc42 splice variant but not Rnd are present. These data indicate that RhoDF or Rnd was lost in either sub-phylum phylum but do not allow inferring which of RhoDF or Rnd emerged first. M. tec. analysis evidenced the presence of the Cdc42 splice variant but failed to identify RhoJQ, RhoDF or Rnd-related peptides among the 106,869 sequences available in the cDNA database. An even more contrasted situation occurred in O. dio., in which we found only RhoABC, Rac and Cdc42 members. This is in consistency with the reduced complexity of this species at the adult stage and the smaller genome size (Seo et al. 2001). In conclusion, our data indicate that the Cdc42 splice variant and three clusters RhoJQ, Rnd and RhoDF emerged in the ancestral chordate, being lost at different extents in urochordates and cephalochordates. Although the close proximity of the branching of Rnd and RhoDF to the RhoABC clade makes it difficult to assess which emerged first (see Figure 1), the recent findings that urochordates might be closer relatives to craniates than cephalochordates (Blair and Hedges 2005; Philippe, Lartillot, and Brinkmann 2005) suggest that the Cdc42 splice variant, RhoJQ and Rnd emerged before RhoDF. Interestingly, we found no Mtl homologue in either chordate species, which indicates that this Rho gene was lost before or early in chordates.

Multiple Rho duplications and emergence of RhoH in Vertebrates

The previous results established that most Rho clusters emerged in chordates, RhoH being the only one missing. All prochordate Rho clusters except Rac and RhoABC are made of unique members whereas two are present in human, which probably reflects the two rounds of whole genome duplication that affected the ancestral vertebrate (2R hypothesis) (Hughes 1999). We examined the fate of duplicated Rho members in the genomes of one sauropsid (Gallus gallus, G.gal), two amphibians (Xenopus tropicalis - X.tro and Xenopus laevis -X.lae.) and three teleost fishes (Takifugu rubripes - F.rub and Tetraodon nigroviridis - T.nig [tetraodontiformes] and Danio rerio - D.rer [cypriniformes]). This panel of vertebrates also includes differentially duplicated genome status, since teleost fishes and Xenopus laevis have encompassed a third duplication (3R) event (Graf and Kobel 1991; Meyer and Van de Peer 2005). Searches in each species (http://www.ensembl.org/) produced many positive hits (10^{-8} cutoff threshold), annotated in the majority as Rho proteins but with many errors due to misplaced exon borders. The distribution of Rho members in these vertebrates is listed Table 3. We identified 19 Rho loci in G. gal., 21 in X. tro., 30 in X. lae., 36 in D. rer., 30 in F. rub. and in T. nig. We found four additional D. rer Rho genes compared to a recent study of a previous genome assembly release (Salas-Vidal et al. 2005). Except RhoH, present in all species as a single member, Rho subgroups contained at least two members in most vertebrate genomes. As expected, additionally duplicated D. rer., F. rub. and T. nig. and X. lae. genomes showed a 1.5- to 2-fold excess of members in most subfamilies, only RhoJ, RhoQ, RhoF and RhoH remaining as single members. In all vertebrate clades, we found orthologues for RhoA, -B and -C, Rac1, -2, -3, RhoG, RhoH, RhoU and -V, RhoBTB1 and -2, and Rnd1, -2 and -3. The absence of Rnd1 in G. gal. and Rac2 in T. rub. needs confirmation since it affects unique genomes and may result from incomplete assemblies. Nevertheless, specific losses were observed that affect two species of a same clade: RhoJ and RhoBTB1, missing in both tetraodontiformes species, and Rnd2, not found in both Xenopus species This suggests that these members were respectively lost in puffer fish and clawed frog lineages. Finally, we found RhoD only in human, which suggests a recent emergence.

Rho members recently emerged in therians and amniotes

The absence of RhoD in vertebrates up to sauropsids prompted us to examine additional species. We found both RhoD and RhoF in placentals Euarchontoglires (mouse and rat, rodents) and Laurasiatheria (dog, carnivore, pig and cow, cetartiodactyles). Analysis of the didelphimorph opossum (Monodelphis domestica, Metatheria) revaled 26 Rho loci, including RhoD and RhoF (Table 3). We next examined

the recently available platypus genome, which belongs to Prototherians, the sibling taxon of therians. We evidenced the presence of four of the five RhoF exons but failed to detect any RhoD related exon sequences (Figure 5A), which strongly suggests that RhoD is present only in therians.

In addition to the classical Rac1 protein, a Rac1b isoform encoded by the same locus was evidenced in tumor cells (Jordan et al. 1999; Schnelzer et al. 2000). Rac1b shows a 19 amino acid extra-domain coded by a short alternative exon located in intron 3, which renders the GTPase constitutively active (Fiegen et al. 2004). To evaluate the physiological importance of Rac1b, we inspected the presence of this alternative exon during evolution. As shown in Figure 5B, we easily detected the exon in all mammals examined including opossum and platypus as well as in chick. The exon was not found in other vertebrates, a feature also associated with a much reduced size of the third intron. This suggests that a specific function associated to Rac1 was gained in amniotes.

Expression of Rho genes in mouse tissues

To compare the ontogeny of the Rho family with physiological functions, we wished to examine the tissue distribution of each member. To this aim, we collected Serial Analysis Gene Expression (SAGE) data from normal mouse tissues. The SAGE method, developed for quantitative analysis of expressed genes (Velculescu et al. 1995), has been widely used to compare mRNA distribution in different tissues or physiological conditions (Harbers and Carninci 2005). Tags corresponding to each Rho member were counted from SAGE libraries derived from 34 tissues. Unique tags were not considered. For each tissue, results are expressed as tags per million in Table 4. RhoA, Rac1 and Cdc42 appear the most ubiquitously expressed Rho members, detected in 97–100% of examined libraries, followed by RhoC, RhoU, RhoB (79–82%), RhoG (74%) and Rac3 (47%). The other members were expressed in 3–29% of libraries only. Several members showed tissue-specific distributions, such as Rac2 and RhoH mostly expressed in hemopoietic tissues, in agreement with their original characterization (Reibel et al. 1991; Dallery-Prudhomme et al. 1997). These data support the notion that the founders RhoA, Rac1, Cdc42 and RhoU are ubiquitously expressed, whereas more recent members evolved toward specific functions. RhoBTB is the only ancient member to display a very narrow expression. This suggests either that this member controls specific events or that it acts in most tissues at very low levels. This might be also the case for RhoV, Rnd1 and Rnd2, counted once and only in a restricted subset of libraries.

Conservation of gene structures, duplications and pseudogenes

In taxons split before vertebrates, we found many cases of specifically duplicated Rac or Rho genes (see Table 1–3). The situation appears more stable in vertebrates, except in the opossum, which showed additional Cdc42, RhoA and RhoG genes (Table 3). As expected, we found supernumerary Rho genes in “3R” genomes (Rac1, RhoG, RhoU, RhoV, RhoA, RhoC, Rnd in X. laevis and bony fishes).

Rho clustering into the eight subclasses shown in Figure 1 was supported by gene structures at least in vertebrates. Members of the Cdc42, Rac and RhoUV subgroups are coded by five/six (see Table S4), six and three exons, respectively, while RhoAC, RhoDF and Rnd members are coded by four, five and five/six exons, respectively. RhoG (Rac subfamily), RhoB (Rho subfamily) and RhoH displayed monoexonic ORFs and likely arose from retrotransposition events. Only in tetraodontiformes, we found variant structures, a three-exon RhoG gene and a four-exon RhoU gene (Table S1), likely pseudogenes since they also have accumulated several frameshift mutations. Of interest, vertebrate gene structures were not fully conserved in chordates. Rac and Rnd in the lancelet are coded by one exon less, while in the sea squirt, Rac and RhoJQ have two exons less, RhoF, one exon less and RhoUV, one exon more. The same situation stands in other coelomates, of which only the sea urchin displayed Rho gene structures similar with vertebrates (Table S1). Since Rho proteins were confidently clustered in all species, this indicates that specific gene rearrangements have occurred in each phylum. This is particularly blatant in the urochordate O. dioca, where genes of the RhoABC subfamily contain four, five or six exons.

Discussion

The goal of the present study was to give an insight into Rho family evolution in eukaryotes. Such analysis had never been done before probably because of the low number of available completed eukaryotic genomes. In this study, we included the most recent genomes such as hemichordates, echinoderms and prototherians to address evolutionary aspects for each Rho subfamily and tentatively correlate these features with physiological traits.

A global evolutionary view of the Rho family is illustrated Figure 6. Our data indicate that Rac is likely the founder member of the family. Rac proteins in the slime mold (mycetozoans) and in plants show physiological roles broader than in fungi/metazoans, in particular control cell polarity and cytokinesis (Rivero and Somesh 2002; Gu, Wang, and Yang 2004). This supports a scenario in which ancestral Rac duplications in fungi/metazoans was associated with early specialization, leading to Cdc42 for the control of cell polarity and Rho for cytokinesis (Pruyne et al. 2004; Jaffe and Hall 2005). Like in trypanosome (Field 2005), the absence of genuine Rho genes in plasmodium or diatom is surprising and raises important issues on which actors substitute, in particular for the control of cell polarity and cytokinesis.

Rho, Cdc42 and RhoBTB emerged from Rac within a 100–200 MY period of time (Hedges et al. 2004). Rho and RhoBTB both branched close to the root of the family, in contrast to Cdc42 confidently related to Rac, which leaves open the possibility that Rho and RhoBTB emerged before metazoans and were lost in early taxons. From bilaterians up to now (i.e. a 1,300 MYA period), Rho, Rac, Cdc42 and RhoBTB were maintained in all animal species, only exceptions being RhoBTB absent in C. elegans and O. dioica. This confirms the well documented roles of Rho, Rac and Cdc42 in basic cell metabolism and lends support to recent data implicating RhoBTB2 (also termed as Deleted in Breast Cancer, DBC2) in the control of proliferation, apoptosis and membrane trafficking (Aspenstrom, Fransson, and Saras 2004; Siripurapu et al. 2005). Two additional members emerged in coelomates (1,100–1,300 MYA ago): Mtl, lost between echninoderms and chordates, and RhoUV, found in all taxons thereafter. First identified and named as Cdc42 related proteins (Aronheim et al. 1998; Tao et al. 2001), RhoUV branched at the vicinity of the Rac/Cdc42 split, in agreement with recent reports (Colicelli 2004; Wennerberg and Der 2004). Despite lack of information on their cellular functions, the presence of RhoUV in early coelomates and the Wnt dependence of RhoU expression (Logan and Nusse 2004) calls for roles in developmental processes.

Three new members delineating two new subfamilies emerged in protochordates (urochordates and cephalochordates), namely RhoJQ, RhoDF and Rnd. RhoJQ derived from Cdc42 and is present in both protochordates. In vertebrates, RhoQ (TC10) and RhoJ (TCL) are prominently expressed in muscle (Murphy et al. 1999; Vignal et al. 2000) and have been implicated in vesicle trafficking (de Toledo et al. 2003) and in insulin-stimulated glucose transport through the Glut-4 transporter (Chang, Chiang, and Saltiel 2005). However, the role of RhoJQ needs to be specified since the control of glucose uptake by insulin and Glut-4 is conserved in chordates and also in drosophila (Escher and Rasmuson-Lestander 1999), which lacks RhoJQ homologue. Interestingly, a recent analysis of 146 nuclear genes supports the grouping of urochordates with vertebrates and that of cephalochordates with echninoderms (Delsuc et al. 2006). If the distribution of Mtl, Rnd, RhoDF and RhoJQ in these taxons equally supports the prior splitting of either urochordates or cephalochordates with respect to vertebrates, it rejects the grouping of cephalochordates and echinoderms, since it would involve an unreasonably high occurrence of homoplasic events. In addition to the three new Rho clusters, a Cdc42 variant appeared in chordates, resulting from alternative splicing of the duplicated 3′ last exon encoding the 29 carboxy-terminal aminoacids of the protein (see Table S4). In mice, and probably in other vertebrates, the new Cdc42b isoform is expressed only in brain, whereas the other (Cdc42u) is expressed ubiquitously (Marks and Kwiatkowski 1996). Both isoforms differ by the nine last amino acids only. Cdc42u and Cdc42b have specific functions since Cdc42u but not Cdc42b contains a dilysine motif critical for binding to the coatomer complex (COP) in the endoplasmic reticulum and shown necessary to induce malignant transformation (Wu et al. 2000). The dilysine motif is present in all eukaryotes down to yeast except in the lancelet. This strengthens the physiological importance of this motif and suggests that an additional exon encoding the dilysine probably exists in the lancelet but was missed in the analysis. The specific function of the second variant in prochordates remains totally obscure in absence of data on its tissue distribution.

After the protochordates, all bony vertebrates examined displayed nearly the same Rho repertoires, suggesting that most additional members arose from whole genome duplications that occurred before the cartilaginous/bony vertebrates split (Panopoulou and Poustka 2005). Availability of lamprey and hagfish genomes will help to elucidate this issue. Our preliminary analysis on limited data sets identified only RhoA, Rac1, Cdc42 and RhoG in Eptatretus (hagfish) and Petromyzon (lamprey) (not shown). RhoH, RhoD and Rac1b showed distinctive behaviors: RhoH, absent in protochordates, is present as a single copy in all vertebrates, indicating that it likely arose after the major duplications or was rapidly lost thereafter. RhoH ontogeny remains obscure, since although found in vertebrates only, its branching is very close to the Rho family root. Hypotheses that RhoH branching is a consequence of sequence shuffling with other Rho members or genuine early divergence are inconclusive. More compelling is the possibility that RhoH derived from distant species and was gained by horizontal transfer, transmitted by either parasites or retrovirus, what would explain its intronless gene structure. This hypothesis is supported by RhoH specific expression in the immune system and its ability to negatively modulate other Rho GTPases (Li et al. 2002), a classical property shared by many pathogen toxins (Aktories and Barbieri 2005).

RhoD showed also a taxon distribution discrepant with its phylogenetic position, only found in therians whereas it apparently duplicated from the RhoDF ancestor in early bony vertebrates. The higher number of paralogous genes in syntheny with RhoD and RhoF (10 vs 6 for RhoA/RhoC and 3 for RhoJ/RhoQ, see http://wolfe.gen.tcd.ie/dup) supports a recent duplication, while the comparison of the ratio of non-synonymous to synonymous substitution rates (4.4 fold higher for RhoD vs RhoF, see Table S5 in supplemental data) suggests that although under selective pressure, RhoD has evolved faster than RhoF. Altogether, these data support the hypothesis that the RhoD/RhoF duplication took place in therians, i.e. 175–220 MYA ago. In cultured cells, RhoD controls endosome dynamics and axon guidance by modulating Src kinase and DIAPH2 formin activities and Semaphorin/Plexin signaling, respectively, all highly conserved in vertebrates (Zanata et al. 2002; Gasman, Kalaidzidis, and Zerial 2003). Therian-limited RhoD expression does not reflect such basic cellular functions and since most studies did not address RhoD specificity versus RhoF, the possibility remains that most functions ascribed to RhoD are actually fulfilled by its closest relative RhoF.

Finally, the minor Rac1b isoform was found exclusively in amniotes. Rac1b protein shows enhanced activity due to a 19 amino acid insertion encoded by an alternative 57 bp exon buried in the third intron (Jordan et al. 1999; Matos, Collard, and Jordan 2003; Fiegen et al. 2004). The 19 aa insert is extremely well conserved and was probably gained upon sequence insertion, since the third Rac1 intron is much shorter in fish and xenopus. Conservation of this alternative exon indicates that Rac1b was positively selected and calls for specific physiological function, possibly in relation with cell adhesion (Chartier et al. 2006).

Comparison of Rho mRNA expression patterns in mouse tissues showed that most members emerged in chordates have a distribution narrower than that of ancient members such as Rho, Rac, Cdc42 and RhoU. This suggests that these latter have basic cellular roles, a notion supported by the early lethality of Rac1- and Cdc42-deficient embryos (Sugihara et al. 1998; Chen et al. 2000). Besides, Rac3, RhoB, RhoC and RhoG, also widely expressed in mice tissues, induce limited defects in the adult but are all dispensable for embryogenesis and post-natal development (Liu et al. 2001; Vigorito et al. 2004; Cho et al. 2005; Corbetta et al. 2005; Hakem et al. 2005). Despite their broad distribution, these members thus seem required only for a narrow range of physiological functions. The current pattern of Rho-deficiency phenotypes actually fits a model in which only one member of each subfamily is critical for embryonic development. One can predict that deficiency in at least one member of RhoUV and RhoBTB subfamilies could also induce severe defects, whereas deficiencies in Rnd, RhoDF and RhoJQ, which delineate the most recent subfamilies, would induce intermediate phenotypes.

A general feature of the Rho family is its high dynamics, illustrated by the high incidence of gain and loss of members along evolution. For instance, the absence of Rac in the yeast S. cerevisiae or S. pombe results from a specific loss since Rac was detected in several other basidiomycetes and ascomycetes. More recently, RhoJ and RhoBTB1 were lost in tetraodontiformes, Rnd2 in Xenopus and Mtl in chordates. If lack of knowledge on the physiological roles of RhoJ, RhoBTB and Rnd makes it difficult to evaluate the impact of their loss, literature is more documented for Mtl/Mig2. In drosophila and nematode, Mtl and its orthologue Mig2 participate with Rac in the control of axon outgrowth and guidance (Zipkin, Kindt, and Kenyon 1997; Lundquist et al. 2001; Hakeda-Suzuki et al. 2002; Ng et al. 2002). The absence of Mtl in chordates suggests either that a particular physiological function was lost or to the contrary, that another Rho-controlled pathway was used to fulfill the same functions as Mtl. It is noteworthy that Mtl loss was paralleled by the emergence of RhoF, RhoJ and Rnd2 in chordates, the latter two being implicated in neurite outgrowth and branching (Fujita et al. 2002; Abe et al. 2003). Expression data (Table 4) suggest that RhoF might be the best candidate. Another example of Rho gene loss is illustrated by urochordates, in which the larvacean Oikopleura dioica encompassed a dramatic reduction in its Rho repertoire (see Table 3). O. dioica is a free-living planktonic organism, which keeps larva morphology and tiny size (<0.5 mm) all along its lifetime. By comparison, ascidians undergo a massive metamorphosis leading to the loss of vertebrate features and growth of specialized organs and tissues. Rac, Cdc42 and Rho proteins are thus sufficient for O. dioica development up to the tailbud stage.

This suggests that these basic GTPases may also be sufficient in ascidian to allow development up to the same stage, the other Rho members being involved in and after metamorphosis, a process which involves intricate patterns of cell proliferation and apoptosis (Chambon et al. 2002; Tarallo and Sordino 2004) and a complete rearrangement of organs (Jeffery and Swalla 1997).

In conclusion, we reported here an exhaustive analysis of the Rho family of GTPases during evolution of eukaryotes, from unicellular organisms of the eukaryotic crown to mammals. We established that the human family contains 20 proteins, MIRO proteins best being considered as a distinct Ras-like subfamily, also conserved in most eukaryotes. Rho members originated from an ancestral Rac and distributed into eight subfamilies, of which four were already present in bilaterians and five in ecdysozoans, two appeared in chordate and the last one in vertebrates. Knowledge of the period at which each subfamily and member appeared, in particular between chordates and vertebrates, combined with comparative embryology and physiology should help specify their functions.

Supplementary material

Aminoacid sequences and features, web databases and murine SAGE data used in the manuscript are listed in supplementary files

Ackowledgements:

We thank G. Bompard and A. Blangy for critical reading of the manuscript and helpful discussions. This work was funded by institutional grants (CNRS), the Association pour la Recherche contre le Cancer (ARC n°3753) and from the GSO Canceropole.

References:

- Abe T , Kato M , Miki H , Takenawa T , Endo T 2003; Small GTPase Tc10 and its homologue RhoT induce N-WASP-mediated long process formation and neurite outgrowth. Journal of Cell Science. 116 :155 -168
- Aktories K , Barbieri JT 2005; Bacterial cytotoxins: targeting eukaryotic switches. Nat Rev Microbiol. 3 :397 -410
- Altschul SF , Madden TL , Schaffer AA , Zhang J , Zhang Z , Miller W , Lipman DJ 1997; Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25 :3389 -3402
- Andersen PR , Devare SG , Tronick SR , Ellis RW , Aaronson SA , Scolnick EM 1981; Generation of BALB-MuSV and Ha-MuSC by type C virus transduction of homologous transforming genes from different species. Cell. 26 :129 -134
- Aronheim A , Broder YC , Cohen A , Fritsch A , Belisle B , Abo A 1998; Chp, a homologue of the GTPase Cdc42Hs, activates the JNK pathway and is implicated in reorganizing the actin cytoskeleton. Current Biology. 8 :1125 -1128
- Aspenstrom P , Fransson A , Saras J 2004; Rho GTPases have diverse effects on the organization of the actin filament system. Biochemical Journal. 377 :327 -337
- Berzat AC , Buss JE , Chenette EJ , Weinbaum CA , Shutes A , Der CJ , Minden A , Cox AD 2005; Transforming activity of the Rho family GTPase, Wrch-1, a Wnt-regulated Cdc42 homolog, is dependent on a novel carboxyl-terminal palmitoylation motif. Journal of Biological Chemistry. 280 :33055 -33065
- Blair JE , Hedges SB 2005; Molecular phylogeny and divergence times of deuterostome animals. Molecular Biology and Evolution. 22 :2275 -2284
- Boon K , Osorio EC , Greenhut SF 2002; An anatomy of normal and malignant gene expression. Proceedings of the National Academy of Sciences of the United States of America. 99 :11287 -11292
- Bourne HR , Sanders DA , McCormick F 1991; The GTPase superfamily: conserved structure and molecular mechanism. Nature. 349 :117 -127
- Bucci C , Chiariello M 2006; Signal transduction gRABs attention. Cellular Signalling. 18 :1 -8
- Chambon JP , Soule J , Pomies P , Fort P , Sahuquet A , Alexandre D , Mangeat PH , Baghdiguian S 2002; Tail regression in Ciona intestinalis (Prochordate) involves a Caspase-dependent apoptosis event associated with ERK activation. Development. 129 :3105 -3114
- Chang L , Chiang SH , Saltiel AR 2005; Insulin Signaling and the Regulation of Glucose Transport. Molecular Medicine.
- Chartier NT , Laine M , Gout S , Pawlak G , Marie CA , Matos P , Block MR , Jacquier-Sarlin MR 2006; Laminin-5-integrin interaction signals through PI 3-kinase and Rac1b to promote assembly of adherens junctions in HT-29 cells. Journal of Cell Science. 119 :31 -46
- Chen F , Ma L , Parrini MC 2000; Cdc42 is required for PIP(2)-induced actin polymerization and early development but not for cell viability. Current Biology. 10 :758 -765
- Cho YJ , Zhang B , Kaartinen V , Haataja L , de Curtis I , Groffen J , Heisterkamp N 2005; Generation of rac3 null mutant mice: role of Rac3 in Bcr/Abl-caused lymphoblastic leukemia. Molecular and Cellular Biology. 25 :5777 -5785
- Colicelli J 2004; Human RAS superfamily proteins and related GTPases. Sci STKE2004. :RE13
- Corbetta S , Gualdoni S , Albertinazzi C , Paris S , Croci L , Consalez GG , de Curtis I 2005; Generation and characterization of Rac3 knockout mice. Molecular and Cellular Biology. 25 :5763 -5776
- Dallery-Prudhomme E , Roumier C , Denis C , Preudhomme C , Kerckaert JP , Galiegue-Zouitina S 1997; Genomic structure and assignment of the RhoH/TTF small GTPase gene (ARHH) to 4p13 by in situ hybridization. Genomics. 43 :89 -94
- de Toledo M , Senic-Matuglia F , Salamero J , Uze G , Comunale F , Fort P , Blangy A 2003; The GTP/GDP cycling of rho GTPase TCL is an essential regulator of the early endocytic pathway. Molecular Biology of the Cell. 14 :4846 -4856
- Delsuc F , Brinkmann H , Chourrout D , Philippe H 2006; Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature. 439 :965 -968
- Donaldson JG , Honda A 2005; Localization and function of Arf family GTPases. Biochemical Society Transactions. 33 :639 -642
- Escher SA , Rasmuson-Lestander A 1999; The Drosophila glucose transporter gene: cDNA sequence, phylogenetic comparisons, analysis of functional sites and secondary structures . Hereditas. 130 :95 -103
- Felsenstein J 1996; Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods in Enzymology. 266 :418 -427
- Fiegen D , Haeusler LC , Blumenstein L , Herbrand U , Dvorsky R , Vetter IR , Ahmadian MR 2004; Alternative splicing of Rac1 generates Rac1b, a self-activating GTPase. Journal of Biological Chemistry. 279 :4743 -4749
- Field MC 2005; Signalling the genome: the Ras-like small GTPase family of trypanosomatids. Trends Parasitol. 21 :447 -450
- Freeman JL , Abo A , Lambeth JD 1996; Rac “insert region” is a novel effector region that is implicated in the activation of NADPH oxidase, but not PAK65. Journal of Biological Chemistry. 271 :19794 -19801
- Fujita H , Katoh H , Ishikawa Y , Mori K , Negishi M 2002; Rapostlin is a novel effector of Rnd2 GTPase inducing neurite branching. Journal of Biological Chemistry. 277 :45428 -45434
- Gasman S , Kalaidzidis Y , Zerial M 2003; RhoD regulates endosome dynamics through Diaphanous-related Formin and Src tyrosine kinase. Nat Cell Biol. 5 :195 -204
- Graf JD , Kobel HR 1991; Genetics of Xenopus laevis. Methods in Cell Biology. 36 :19 -34
- Gu Y , Wang Z , Yang Z 2004; ROP/RAC GTPase: an old new master regulator for plant signaling. Current Opinion in Plant Biology. 7 :527 -536
- Hakeda-Suzuki S , Ng J , Tzu J , Dietzl G , Sun Y , Harms M , Nardine T , Luo L , Dickson BJ 2002; Rac function and regulation during Drosophila development. Nature. 416 :438 -442

- Hakem A , Sanchez-Sweatman O , You-Ten A , Duncan G , Wakeham A , Khokha R , Mak TW 2005; RhoC is dispensable for embryogenesis and tumor initiation but essential for metastasis. Genes and Development. 19 :1974 -1979
- Hall A 1998; Rho GTPases and the actin cytoskeleton. Science. 279 :509 -514
- Hancock JF , Parton RG 2005; Ras plasma membrane signalling platforms. Biochemical Journal. 389 :1 -11
- Harbers M , Carninci P 2005; Tag-based approaches for transcriptome research and genome annotation. Nat Methods. 2 :495 -502
- Hedges SB , Blair JE , Venturi ML , Shoe JL 2004; A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evol Biol. 4 :2
- Hughes AL 1999; Phylogenies of developmentally important proteins do not support the hypothesis of two rounds of genome duplication early in vertebrate history. Journal of Molecular Evolution. 48 :565 -576
- Jaffe AB , Hall A 2005; Rho GTPases: biochemistry and biology. Annual Review of Cell and Developmental Biology. 21 :247 -269
- Jeanmougin F , Thompson JD , Gouy M , Higgins DG , Gibson TJ 1998; Multiple sequence alignment with Clustal X. Trends in Biochemical Sciences. 23 :403 -405
- Jeffery WR , Swalla BJ Editor: Gilbert S 1997; Embryology of the Tunicates. :331 -364 Embryology: Constructing the organisms. Sinauer; Sunderland
- Jordan P , Brazao R , Boavida MG , Gespach C , Chastre E 1999; Cloning of a novel human Rac1b splice variant with increased expression in colorectal tumors. Oncogene. 18 : 6835 -6839
- Li X , Bu X , Lu B , Avraham H , Flavell RA , Lim B 2002; The hematopoiesis-sspecific GTP-binding protein RhoH is GTPase deficient and modulates activities of specific other Rho GTPases by an inhibitory function. Molecular and Cellular Biology. 22 :1158 -1171
- Liu AX , Rane N , Liu JP , Prendergast GC 2001; RhoB is dispensable for mouse development, but it modifies susceptibility to tumor formation as well as cell adhesion and growth factor signaling in transformed cells. Molecular and Cellular Biology. 21 :6906 -6912
- Logan CY , Nusse R 2004; The Wnt signaling pathway in development and disease. Annual Review of Cell and Developmental Biology. 20 :781 -810
- Lundquist EA , Reddien PW , Hartwieg E , Horvitz HR , Bargmann CI 2001; Three C. elegans Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. Development. 128 :4475 -4488
- Madaule P , Axel R 1985; A novel ras-related gene family. Cell. 41 :31 -40
- Marks PW , Kwiatkowski DJ 1996; Genomic organization and chromosomal location of murine Cdc42. Genomics. 38 :13 -18
- Matos P , Collard JG , Jordan P 2003; Tumor-related alternatively spliced Rac1b is not regulated by Rho-GDP dissociation inhibitors and exhibits selective downstream signaling. Journal of Biological Chemistry. 278 :50442 -50448
- Meyer A , Van de Peer Y 2005; From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). Bioessays. 27 :937 -945
- Murphy GA , Solski PA , Jillian SA , Perez de la Ossa P , D'Eustachio P , Der CJ , Rush MG 1999; Cellular functions of TC10, a Rho family GTPase: regulation of morphology, signal transduction and cell growth. Oncogene. 18 :3831 -3845
- Ng J , Nardine T , Harms M , Tzu J , Goldstein A , Sun Y , Dietzl G , Dickson BJ , Luo L 2002; Rac GTPases control axon growth, guidance and branching. Nature. 416 :442 -447
- Page RD 1996; TreeView: an application to display phylogenetic trees on personal computers. Computer Applications in the Biosciences. 12 :357 -358
- Panopoulou G , Poustka AJ 2005; Timing and mechanism of ancient vertebrate genome duplications -- the adventure of a hypothesis. Trends in Genetics. 21 :559 -567
- Pemberton LF , Paschal BM 2005; Mechanisms of receptor-mediated nuclear import and nuclear export. Traffic. 6 :187 -198
- Philippe H , Lartillot N , Brinkmann H 2005; Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Molecular Biology and Evolution. 22 :1246 -1253
- Philips A , Blein M , Robert A , Chambon JP , Baghdiguian S , Weill M , Fort P 2003; Ascidians as a vertebrate-like model organism for physiological studies of Rho GTPase signaling. Biologie Cellulaire. 95 :295 -302
- Pruyne D , Legesse-Miller A , Gao L , Dong Y , Bretscher A 2004; Mechanisms of polarized growth and organelle segregation in yeast. Annual Review of Cell and Developmental Biology. 20 :559 -591
- Reibel L , Dorseuil O , Stancou R , Bertoglio J , Gacon G 1991; A hemopoietic specific gene encoding a small GTP binding protein is overexpressed during T cell activation. Biochemical and Biophysical Research Communications. 175 :451 -458
- Ridley AJ 2001; Rho family proteins: coordinating cell responses. Trends in Cell Biology. 11 :471 -477
- Rivero F , Dislich H , Glockner G , Noegel AA 2001; The Dictyostelium discoideum family of Rho-related proteins. Nucleic Acids Res. 29 :1068 -1079
- Rivero F , Somesh BP 2002; Signal transduction pathways regulated by Rho GTPases in Dictyostelium. Journal of Muscle Research and Cell Motility. 23 :737 -749
- Rozas J , Sanchez-DelBarrio JC , Messeguer X , Rozas R 2003; DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19 :2496 -2497
- Saitou N , Nei M 1987; The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution. 4 :406 -425
- Salas-Vidal E , Meijer AH , Cheng X , Spaink HP 2005; Genomic annotation and expression analysis of the zebrafish Rho small GTPase family during development and bacterial infection. Genomics. 86 :25 -37
- Schnelzer A , Prechtel D , Knaus U , Dehne K , Gerhard M , Graeff H , Harbeck N , Schmitt M , Lengyel E 2000; Rac1 in human breast cancer: overexpression, mutation analysis, and characterization of a new isoform, Rac1b. Oncogene. 19 :3013 -3020
- Seeburg PH , Colby WW , Capon DJ , Goeddel DV , Levinson AD 1984; Biological properties of human c-Ha-ras1 genes mutated at codon 12. Nature. 312 :71 -75
- Seo HC , Kube M , Edvardsen RB 2001; Miniature genome in the marine chordate Oikopleura dioica. Science. 294 :2506
- Siripurapu V , Meth J , Kobayashi N , Hamaguchi M 2005; DBC2 significantly influences cell-cycle, apoptosis, cytoskeleton and membrane-trafficking pathways. Journal of Molecular Biology. 346 :83 -89
- Sugihara K , Nakatsuji N , Nakamura K 1998; Rac1 is required for the formation of three germ layers during gastrulation. Oncogene. 17 :3427 -3433
- Tanaka K , Takai Y 1998; Control of reorganization of the actin cytoskeleton by Rho family small GTP-binding proteins in yeast. Current Opinion in Cell Biology. 10 :112 -116
- Tao W , Pennica D , Xu L , Kalejta RF , Levine AJ 2001; Wrch-1, a novel member of the Rho gene family that is regulated by Wnt-1. Genes and Development. 15 :1796 -1807
- Tarallo R , Sordino P 2004; Time course of programmed cell death in Ciona intestinalis in relation to mitotic activity and MAPK signaling. Developmental Dynamics. 230 :251 -262
- Valster AH , Hepler PK , Chernoff J 2000; Plant GTPases: the Rhos in bloom. Trends in Cell Biology. 10 :141 -146
- Velculescu VE , Zhang L , Vogelstein B , Kinzler KW 1995; Serial analysis of gene expression. Science. 270 :484 -487
- Venter JC , Adams MD , Myers EW 2001; The sequence of the human genome. Science. 291 :1304 -1351
- Vignal E , De Toledo M , Comunale F , Ladopoulou A , Gauthier-Rouviere C , Blangy A , Fort P 2000; Characterization of TCL, a new GTPase of the rho family related to TC10 and Ccdc42. Journal of Biological Chemistry. 275 :36457 -36464
- Vigorito E , Bell S , Hebeis BJ , Reynolds H , McAdam S , Emson PC , McKenzie A , Turner M 2004; Immunological function in mice lacking the Rac-related GTPase RhoG. Molecular and Cellular Biology. 24 :719 -729
- Wennerberg K , Der CJ 2004; Rho-family GTPases: it's not only Rac and Rho (and I like it). Journal of Cell Science. 117 :1301 -1312
- Wennerberg K , Rossman KL , Der CJ 2005; The Ras superfamily at a glance. Journal of Cell Science. 118 :843 -846
- Wherlock M , Mellor H 2002; The Rho GTPase family: a Racs to Wrchs story. Journal of Cell Science. 115 :239 -240
- Wu WJ , Erickson JW , Lin R , Cerione RA 2000; The gamma-subunit of the coatomer complex binds Cdc42 to mediate transformation. Nature. 405 :800 -804
- Zanata SM , Hovatta I , Rohm B , Puschel AW 2002; Antagonistic effects of Rnd1 and RhoD GTPases regulate receptor activity in Semaphorin 3A-induced cytoskeletal collapse. Journal of Neuroscience. 22 :471 -477
- Zipkin ID , Kindt RM , Kenyon CJ 1997; Role of a new Rho family member in cell migration and axon guidance in C. elegans. Cell. 90 :883 -894

Figure 1

Delineation and structure of the human Rho family

Proteins considered so far as Rho members were aligned with GTPases of other Ras-like families and the unrooted tree was obtained by NJ (ClustalX). Bootstrap values at critical nodes show that MIRO proteins constitute a distinct Ras-like family and RhoBTB3 is branched outside the Rho family. Identical topology was obtained using maximum likelihood (ProML3.6.3). Only the Rho domains, corresponding to aminoacids 5-173 of Rac1, were used for the alignment. Structuration into 4 clusters and 8 sub-families is figured by light and dark grey ellipses respectively. When different, common names are figured into brackets under the HUGO nomenclature.

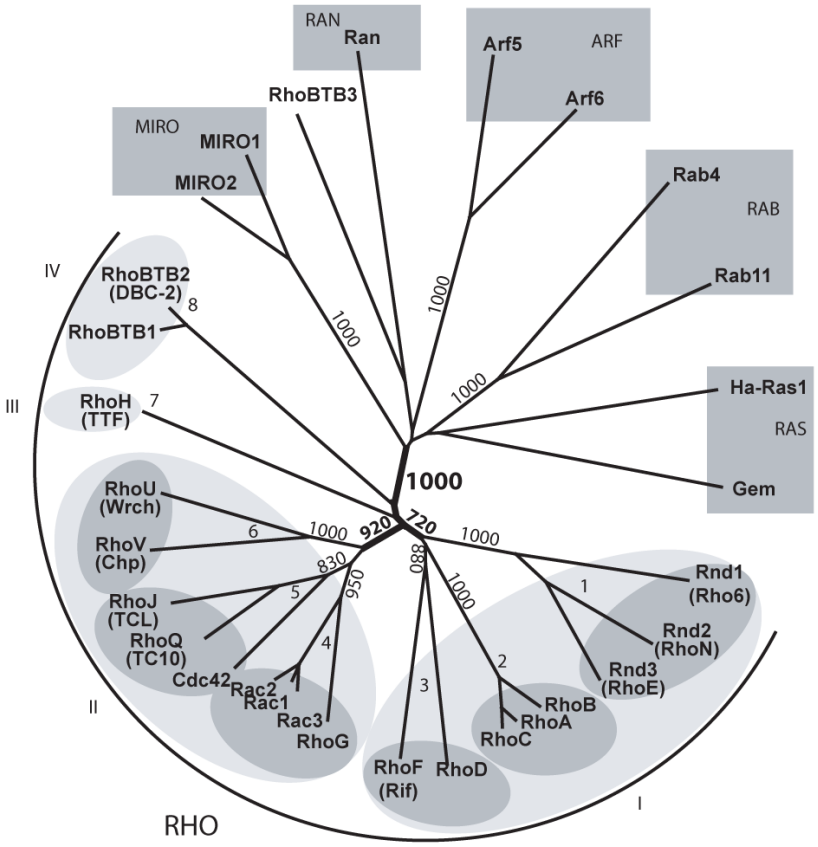


Figure 2

Rac as the founder of the Rho family

Rho sequences from fungi (Saccharomyces cerevisiae - Sc, Yarrowia lipolytica - Yl), entamoeba (Entamoeba histolytica - Eh), mycetozoans (Dictyostelium discoideum - Dd), alveolates (Tetrahymena thermophila - Tt), stramenopiles (Phytophthora ramorum - Pr) and plants (Arabidopsis thaliana - At) were aligned using ClustalX. Hydra magnipapillata (Hm) sequences were included as metazoan Rho sequences and Rab sequences as an external group. Only bootstrap values >700 are indicated on the NJ tree.

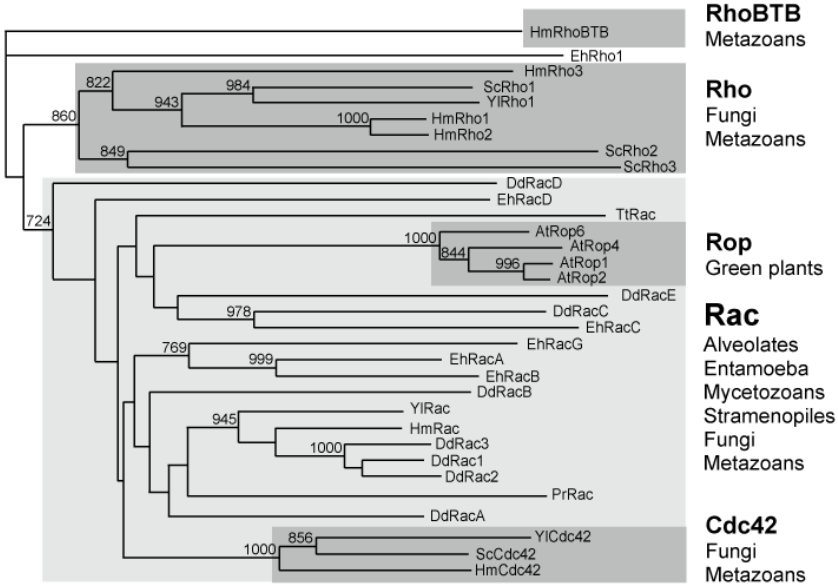
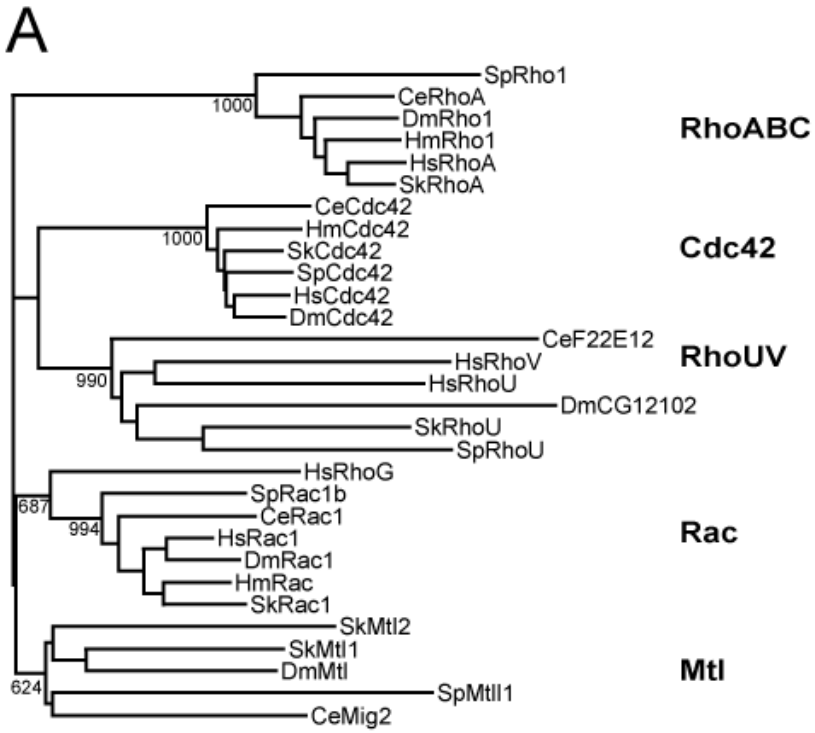


Figure 3

Five Rho subfamilies in Coelomates

A: Rho sequences from Drosophila melanogaster (Dm), Caenorhabditis elegans (Ce), Sacchoglossus kowalevskii (Sk), Strongylocentrotus purpuratus (Sp) were aligned with ClustalX. Hydra magnipapillata (Hm) and human (Hs) sequences were included as acoelomate and chordate groups. Only bootstrap values >600 are indicated on the NJ tree. **B:** The amino acid sequences of RhoUV members were aligned with ClustalX. Human RhoA, Rac1 and Cdc42 were included as outgoups to delineate residues specific of the RhoUV subfamily (grey shaded). CeF22E12 (CeRhoU) apomorphic positions are in bold.



B

```
                6
HsRhoA          KKLVIVGDGACGKTCLLIVFSKDQFPEVYVPTVFENYVADIEVDGKQVELALWDTAGQEDY
HsRac1          IKCVVVGDGAVGKTCLLISYTTNAFPGEYIPTVFDNYSANVMVDGKPVNLGLWDTAGQEDY
HsCdc42         IKCVVVGDGAVGKTCLLISYTTNKFPSEYVPTVFDNYAVTVMIGGEPYTLGLFDTAGQEDY
HsRhoV          IKCVLVGDGAVGKSSLIVSYTCNGYPARYRPTALDTFSVQVLVDGAPVRIELWDTAGQEDF
HsRhoU          VKCVLVGDGAVGKTSLVVSYTTNGYPTEYIPTAFDNFSAVVSVDGRPVRLQLCDTAGQDEF
SkRhoU          VKCVLVGDGAVGKTSLIVSYTTNGYPTEYIPTAYDNYAVVVNVDNKPVRLQLCDTAGQDDF
SpRhoU          LKCVLVGDGAVGKTSLVVSYTTNGYPVQYVPTAFDNYSVVIRVDEKPIKLQICDTAGQDDF
DmCG12102       IKCVLVGDGAVGKTNLILSYLENRFNPEHIPTASDIYNADVNVNESPVHLTICDTAGQDTL
CiWrch          LKCVLLGDGAVGKTSLVVSYSTNGYPTEYVPTAFDNYSVRIEIDGNALRLQICDTAGQDEF
CeF22E12        LKCIFVGNAAVGKTSMIVSYTTNVYPHNYVPTAFDNFSVVVLVDKKPIRLQLWDTAGQSSF

                67
HsRhoA          DRLRPLSYPDTDVILMCFSIDSPDSLENIPEKWTPEVKHFCPNVPIILVGNKKDLRNDEHT
HsRac1          DRLRPLSYPQTDVFLICFSLVSPASFENVRAKWYPEVRHHCPNTPIILVGTKLDLRDDKDT
HsCdc42         DRLRPLSYPQTDVFLVCFSVVSPSSFENVKEKWVPEITHHCPKTPFLLVGTQIDLRDDPST
HsRhoV          DRLRSLCYPDTDVFLACFSVVQPSSFQNITEKWLPEIRTHNPQAPVLLVGTQADLRDDVNV
HsRhoU          DKLRPLCYTNTDIFLLCFSVVSPSSFQNVSEKWVPEIRCHCPKAPIILVGTQSDLREDVKV
SkRhoU          DSLRPLCYPQTDVFLVCFSVVSPTSFHNVVEKWVPEVRKHNKKTPILLVGTQCDLGNDVKV
SpRhoU          DSLRPLCYPQTDVFLLCFSVVSPTSFHNVLEKWLPEIRRHNPKTPILLIGTQCDLRTDVNV
DmCG12102       DPLRELCYPDSDVFLLCFSVVKPETFRAIKTKWAPKFAK--TKAALILVGTQADLRTSPNV
CiWrch          DGMRPLCYPGTDVVLVCFSVVRPTSLCNIRDKWLPEIKKYLPKVPVVLVGTQTDLRTSLDV
CeF22E12        DTLRPLCYTDADVFVIVYSVVDLQSFEDVSHHWYPEVTKRNPGTKLILVGTQADQRWQ---

                128
HsRhoA          RRELAKMKQE-PVKPEEGRDMANRIGAFGYMECSAKTKDGVREVFEMATRAALQ-------
HsRac1          IEKLKEKKLT-PITYPQGLAMAKEIGAVKYLECSALTQRGLKTVFDEAIRAVLCPP-----
HsCdc42         IEKLAKNKQK-PITPETAEKLARDLKAVKYVECSALTQKGLKNVFDEAILAALEPP-----
HsRhoV          LIQLDQGGREGPVPQPQAQGLAEKIRACCYLECSALTQKNLKEVFDSAILSAIEHK-----
HsRhoU          LIELDKCK-EKPVPEEAAKLCAEEIKAASYIECSALTQKNLKEVFDAAIVAGIQYS-----
SkRhoU          PIDLAKF-HERLSLKEDARSCAARISAISY-ECSALTQKNIEKVF
SpRhoU          LIDLSKH-EERPICEEEASMRASRINAVGYIECSALTQHNLKEVFDAVILTGLEYTKPPKS
DmCG12102       LNKLQTNGEEA-ISYADAWDLATTIGAK-YIETSSATQDKVKDVFDTAIWEGL--------
CiWrch          LVDLARY-SERPVTTEEAGAFAKQCGACGYVECSALTQKNLKQVFDTAILEGMDF--ADRR
CeF22E12        -------VRGNTVTQLRGKALADRLGAE-FFECSALTQHNLKQMFDAAILAGLEG------

                Rho insert
                181
HsRhoA          ----------------------------ARRGKKKSGCLVL
HsRac1          ----------------------------PVKKRKRKCLLL
HsCdc42         ----------------------------EPKKSRRCVLL
HsRhoV          --------ARLEKKLNAKG-----VRTLSRCRWKKFFCFV-
HsRhoU          --------DTQQQPKKSKSRTPDKMKNLSKSWWKKYCCFV-
SpRhoU          SKSLRRRLSRKEKIPPPPKRAQDTKPKKKGALWKKLCCIPT
DmCG12102       ----------------------VPTTLPPTPSFWRKLFCLA-
CiWrch          QRKSKRKMRSESKRSHSEKSTTPRTLKCKPRAWLRKYCCAA
CeF22E12        ---------KKNREKRSNFYTKSSKIKERVARFITRTRNLI
```

Figure 4

Seven Rho subfamilies in Chordates

Rho sequences from the cephalochordate Branchiostoma floridae (Bf) and from the urochordates Ciona intestinalis (Ci, ascidian) and Oikopleura dioica (Od, larvacean) were aligned with ClustalX. Human Rho sequences were included as vertebrate outgroups. Only bootstrap values >500 are indicated on the NJ tree.

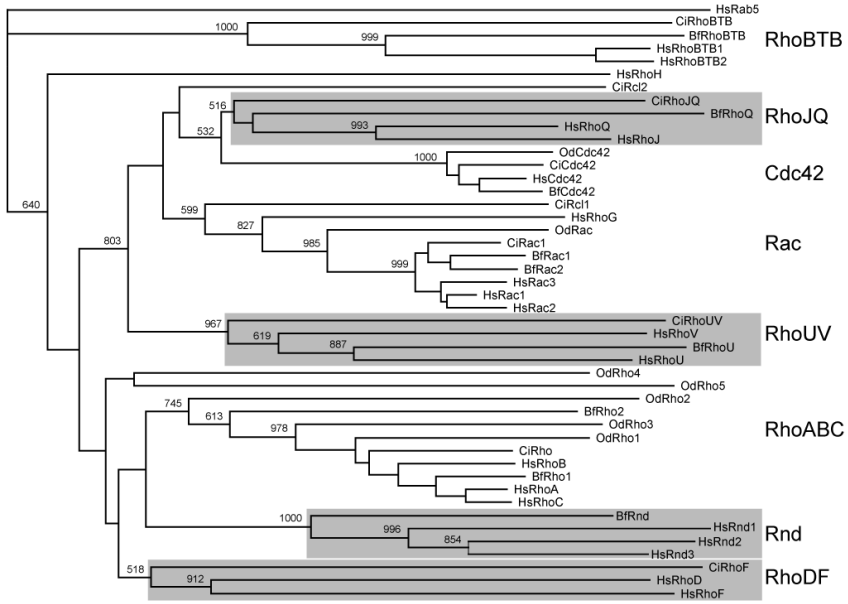


Figure 5

Evolution of Rac1b and RhoD in Vertebrates

A: Vertebrate genomes were searched for the presence of the 57 bp Rac1b-specific exon. For each considered species is shown the predicted peptide, the position of the additional exon upstream of the normal 4th Rac1 exon, and the size of the third exon. **B:** RhoD and RhoF homologues were searched in mouse (Mm), dog (Cf), pig (Ss), opossum (Md), platypus (Oa) and chicken (Gg) and aligned with human sequences using ClustalX. Human Cdc42 and RhoA were included as external outgroups.

A

Species	Rac1b exon sequence	Position 5' from E4	I3 length (kb)
Man	VGETYGKDITSRGKDKPIA	+1.4 kb	8
Chimp	VGETYGKDITSRGKDKPIA	+1.4 kb	8
Mouse	VGDTCGKDRPSRGKDKPIA	+1.5 kb	6.5
Rat	VGDTCGKDRTSRGKDKPIA	+1.6 kb	7
Dog	VGETYGKDIPSRGKDKLIA	+1.6 kb	7.8
Cow	VGETYGKEIPSRGKEKPIA	+1.4 kb	4.7
Armadillo	VGDTYGKDITTRGKDKPIA	+1.5 kb	nd
Elephant	VGDTYGKDSATRGKDKPIA	+1.3 kb	nd
Platypus	VGGTYGKDINSRVKDEPIA	nd	nd
Opossum	VGGTYSKDISSKVKDEPIA	+12.2 kb	14.1
Chick	VGGTNGKNISSSLTDQPIA	+1.9 kb	3.8
Xenopus	not detected	-	0.1
Zebrafish	not detected	-	0.8

B

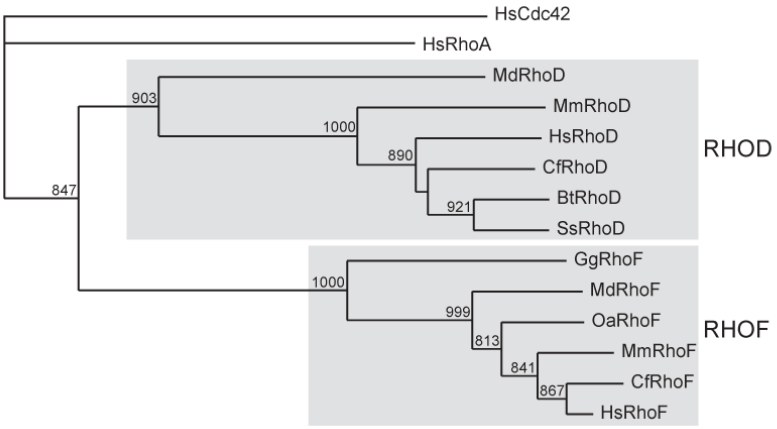


Figure 6

Evolutionary synopsis of the Rho family

The phylogenetic tree of Figure 1 was redrawn taking into accounts the distribution of Rho subfamilies in the examined taxa. Shaded triangles indicate roots and intervals of emergence of the subfamilies. Scale time is in million years (MYA). Broken lines represent discrepancies between inferred phylogeny and observed emergence. † indicates subfamily extinction.

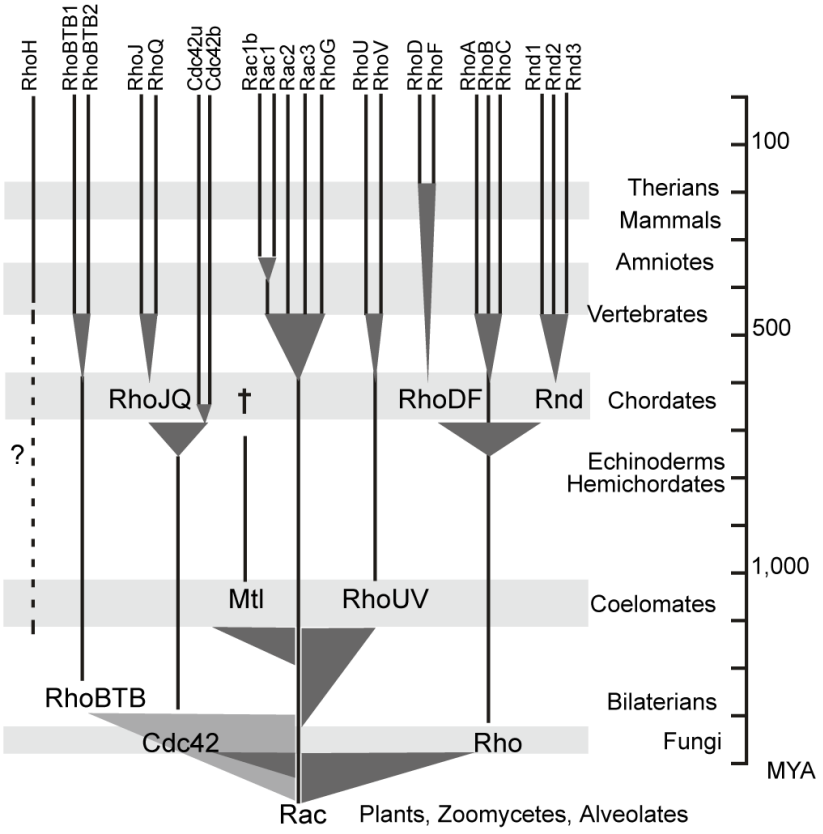


Table 1
Rho subfamilies before Chordates

Taxon	Species	Cdc42	Rac	Rho	RhoBTB	RhoUV
Cnidarians	Hydra magnipapillata	HmCdc42	HmRac	HmRho1 HmRho2 HmRho3	HmRhoBTB	absent
Porifera	Reniera sp. JGI-2005	RCdc42	Rrac1 RRac2 RRac3 RRac4 RRac5	Rrho1	absent	absent
Acoelomates	Schisostoma mansoni	SmCdc42	SmRac1 SmRac2	SmRho1 SmRho2 SmRho3 SmRho4 SmRho5	absent	absent
	Schisostoma japonicum	SjCdc42	SjRac1 SjRac2	SjRho1 SjRho2 SjRho3 SjRho4 SjRho5	absent	absent
Nematodesa	Caenorhabditis elegans	CeCdc42	CeRac1 CeRac2 CeMig2	CeRho	absent	CeRhoU
Insectsb	Drosophila melanogaster	DmCdc42	DmRac1 DmRac2 DmMtl	DmRho1	DmRhoBTB	DmRhoU
Echinoderms	Strongylocentrotus purpuratus	SpCdc42	SpRac1 SpRac2 SpRac3 SpRac4 SpMtl	SpRhol SpRho2 SpRho3	SpRhoBTB	SpRhoU
Hemichordates	Saccoglossus kowalevskii	SkCdc42 SkCdc42p	SkRac1 SkRac2 SkMtl	SkRho1	not foundc	SkRhoU

a Notes Y32F6B.3 was omitted since its Rho membership is uncertain and is restricted to nematodes.
b RhoL was omitted since it lacks the Rho-specific insert and is restricted to insects.

c Members are considered as “absent” when missing in genomic data and only “not found” when missing in EST database.

Table 2
Rho subfamilies in Chordates

	Cephalochordates	**Urochordates**		
		Ascidiacea		**Appendicularia**
Sub-family	Branchiostoma floridae	Enterogona Ciona intestinalis	Stolidobranchia Molgula tectiformis	Oikopleura dioica
Cdc42	BfCdc42	CiCdc42a	MtCdc42a	OdCdc42
RhoJQ	BfRhoJQ	CiRhoJQ	not found	absent
Rac	BfRac1 BfRac2	CiRac1 CiRac2 CiRac3a CiRac3b CiRac4 CiRcl1 CiRcl2	MtRac	OdRac
RhoUV	BfRhoUV	CiRhoUV	MtRhoUV	
RhoABC	BfRho1 BfRho2	CiRho1	MtRho1 MtRho2	OdRho1 OdRho2 OdRho3 OdRho4 OdRho5
RhoDF	absent	CiRhoF	not found	absent
Rnd	BfRnd	absent	not found	absent
RhoBTB	BfRhoBTB	CiRhoBTB	MtRhoBTB	absent

a Note Two Cdc42 isoforms differing in their carboxy terminus were identified (see Table S4)

Table 3
Rho subfamilies in Vertebrates

Sub-family	**Homo sapiens**	**Monodelphis domestica**	**Gallus gallus**	**Xenopus tropicalis**	**Xenopus laevis**	**Brachydanio rerio**	**Tatifugu rubripes**	**Tetraodon nigroviri**
Cdc42a	HsCdc42	MdCdc42a MdCdc42b MdCdc42c	GgCdc42	XtCdc42	XlCdc42	BrCdc42a BrCdc42b BrCdc42c	FrCdc42a1 FrCdc42a2 FrCdc42b	TnCdc42a1 TnCdc42a2 TnCdc42b
	HsRhoJ	MdRhoJ	GgRhoJ	XtRhoJ	XlRhoJ	BrRhoJ	absent	Absent
	HsRhoQ	MdRhoQ	GgRhoQ	XtRhoQ	not found	BrRhoQ	FrRhoQ	TnRhoQ
Rac	HsRac1 HsRac1bb	MdRac1 MdRac1bb	GgRac1 GgRac1bb	XtRac1	XlRac1a XlRac1b XlRac1c1 XlRac1c2	BrRac1a BrRac1b	FrRac1a FrRac1b	TnRac1a TnRac1b TnRac1b1 TnRac1b2

	HsRac2	MdRac2	GgRac2	XtRac2	XlRac2	BrRac2	FrRac2	TnRac2
	HsRac3	MdRac3	GgRac3	XtRac3	not found	BrRac3	FrRac3	Absent
	HsRhoG	MdRhoG	GgRhoG	XtRhoG1	XlRhoG1	BrRhoG1	FrRhoG1	TnRhoG1
		MdRhoG1		XtRhoG2a		BrRhoG2a	FrRhoG2	TnRhoG2
		MdRhoG2		XtRhoG2b		BrRhoG2b		
						BrRhoG3		
						BrRhoG3p		
						BrRhoG4		
RhoUV	HsRhoU	MdRhoU	GgRhoU	XtRhoU	XlRhoU1	BrRhoU1	FrRhoU1	TnRhoU1
					XlRhoU2a	BrRhoU2	FrRhoU2	TnRhoU2
					XlRhoU2b	BrRhoU3	FrRhoU3	
	HsRhoV	MdRhoV	GgRhoV	XtRhoV	XlRhoVl	BrRhoV	FrRhoV1	TnRhoV1
					XlRhoV2		FrRhoV2	TnRhoV2
Rho	HsRhoA	MdRhoA	GgRhoA	XtRhoA1	XlRhoA1a	BrRhoA1	FrRhoA1a	TnRhoA1a
		MdRhoAps1	GgRhoAp	XtRhoA2	XlRhoA1b	BrRhoA2	FrRhoA1b	TnRhoA1b
		MdRhoAps2			XlRhoA1c	BrRhoACa	FrRhoAC	TnRhoAC
					XlRhoA2	BrRhoACb		
	HsRhoB	MdRhoB	GgRhoB	XtRhoB	XlRhoB1	BrRhoB	FrRhoB	TnRhoB
					XlRhoB2			
	HsRhoC	MdRhoC	GgRhoC	XtRhoC	XlRhoC1	BrRhoC1a	FrRhoC1	TnRhoC1a
					XlRhoC2	BrRhoC1b	FrRhoC2	TnRhoC1b
						BrRhoC2		TnRhoC2
RhoBTB	HsRhoBTB1	MdRhoBTB1	GgRhoBTB1	XtRhoBTB1	XlRhoBTB1	BrRhoBTB1	absent	Absent
	HsRhoBTB2	MdRhoBTB2	GgRhoBTB2	XtRhoBTB2	not found	BrRhoBTB2a	FrRhoBTB2a	TnRhoBTB2a
						BrRhoBTB2b		
						BrRhoBTB2c	FrRhoBTB2c	TnRhoBTB2c
RhoDF	HsRhoD	MdRhoD	absent	absent	not found	absent	absent	Absent
	HsRhoF	MdRhoF	GgRhoF	XtRHoF	XlRhoF1	BrRhoF	FrRhoF	TnRhoF
					XlRhoF2			
RhoH	HsRhoH	MdRhoH	GgRhoH	XtRhoH	not found	BrRhoH	FrRhoH	TnRhoH
Rnd	HsRnd1	MdRnd1	absent	XtRnd1	XlRnd1a	BrRnd1a	FrRnd1a	TnRnd1
					XlRnd1b	BrRnd1b	FrRnd1b	
	HsRnd2	MdRnd2	GgRnd2	absent	not found	BrRnd2	FrRnd2a	TnRnd2a
							FrRnd2b	TnRnd2b
	HsRnd3	MdRnd3	GgRnd3	XtRnd3	XlRnd3a	BrRnd3a	FrRnd3	TnRnd3
					XlRnd3b	BrRnd3b		

a Notes Vertebrate Cdc42 members have two isoforms generated by differential splicing (see Table S4)
b Isoform from the same locus as Rac1

Table 4
SAGE analysis of Rho mRNA expression in mouse tissues

Mouse tissue	RhoA	Rac1	Cdc42	RhoC	RhoU	RhoB	RhoG	Rac3	RhoF	Rac2	RhoQ	RhoD	RhoJ	RhoH	Rnd3	RhoBTB1	RhoBTB2	Total
Brain	152	98	89	18		27	89		116								27	11173
Branchial arch	84	1215	76	396	51	110		59										11854
Visual cortex	71	1018	28	50		256		107	28									14048
Cerebellum	298	332	160	69	23	160	34		23									87344
Hypothalamus	53	45	15	15	15	45	30	30							23	15		13286
Adrenal gland	94	529	145	58	65	232			15		15	44	15					13786
Skin	70	35	35		35	52	52											57206
Mammary gland	193	422	96	170	30	118	37				37							13506
Placenta	191	616	21	478	85	234	21											94124
Lung	238	754	92	246	23	62	31	62										13004
Stomach	277	268	157	92	65	55	46				28	18						10828
Small Intestine	142	1101	47	218	19	133		57				19						10534
Large intestine	190	523	86	228	29	86	29	38				19						10511
Pancreas	196	739	309	243	30	65	46				56							10691
Spleen	78	1252	141	282	47	110	31	86		47				59				12778
Thymus	313	627	58	33	16	41	74	33	25	313				487				12122
T-cell	226	103	246				123		123	698				123				48721
Bone marrow	138						276		138	92								21770
Uterus	103	690	103	131	19	56		93				19						10721
Prostate	238	378	227	130	32	119		22			43		22					92631
Kidney	74	385	139	82	16			41										12192
Bladder	235	588	132	406	66	86	29	74	22		29	15	15					13596
Liver	120	251	44	55		33	33											
Heart ventricle	136	336	100	218	64	55	27	18										11401
Heart atrium	192	522	107	117	21	21	43											93835
Skeletal muscle	128	597	111	248	17	68	34			17	26							11716
Hindlimb bud	219	88	205	44	29		102								29			68349
Forelimb bud	190	88	102	88														68302
Ovary	137	561	68	106	15	68	15	15										13180
Testis	117	524	142	50	58	50												12012
ES cells	196	65	65			98	753		65	229								30536
Embryo fibroblasts	213	273	72	319	21	81	51	94	9	21		30	9		9			23482

NOTE. mRNA is expressed as positive tags per million sequenced tags (Total tags). Only tags found at least twice were considered. Unfilled cells indicate too low levels to be estimated.