Genomic annotation and expression analysis of the zebrafish Rho small GTPase family during development and bacterial infection

Salas-Vidal, E.; Meijer, A.H.; Cheng, X.; Spaink, H.P.

Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 86 (2005) 25 – 37

GENOMICS

www.elsevier.com/locate/ygeno

Genomic annotation and expression analysis of the zebrafish Rho small GTPase family during development and bacterial infection☆,☆☆

Enrique Salas-Vidal[a,b,*], Annemarie H. Meijer[a], Xi Cheng[a], Herman P. Spaink[a]

[a]*Institute of Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands*
[b]*Instituto de Biotecnología, UNAM, AP 510-3, Cuernavaca, Morelos 62271, México*

Received 22 December 2004; accepted 24 March 2005
Available online 26 April 2005

Abstract

The zebrafish genomic sequence database was analyzed for the presence of genes encoding members of the Rho small GTPases. The analysis shows the presence of 32 zebrafish Rho genes representing one or more homologs of the human *RHOA, RND3, RHOF, RHOG, RHOH, RHOJ, RHOU, RHOV, CDC42, RAC1, RAC2, RAC3, RND1, RHOBTB1, RHOBTB2, RHOBTB3,* and *RHOT1* genes. By expression analysis using reverse transcriptase-PCR we show that at least 20 of the predicted zebrafish small GTPase genes are expressed in the adult stage. Interestingly, only 5 of these were found to be expressed at early embryonic stages, including *rhoab, rhoad, cdc42a, cdc42c,* and *rac1a*. We observed a strong upregulation of zebrafish *rhogb* expression after *Mycobacterium marinum* infection of adult fish. This complete annotation study provides a firm basis for the use of zebrafish as a model for analysis of Rho GTPase function in vertebrate development and the innate immune system.

Keywords: Embryogenesis; Innate immunity; Mycobacterial disease; *Danio rerio*; Monomeric G proteins

The family of Rho small GTPases comprises a large subgroup of genes of the Ras superfamily of small GTPases. The members of this group are distinguished from other small GTPases because they possess a distinctive insert region in the GTPase domain [1]. Rho GTPases are guanine nucleotide binding proteins that cycle from the active GTP-bound state to the inactive GDP-bound state. GTP binding induces conformational changes in at least two regions of Rho GTPases, called switch 1 and switch 2. These conformational changes increase the binding affinity for target or effector proteins, which in turn stimulate diverse signaling pathways that mediate the different Rho GTPase functions [2].

Rho GTPases are known to play key roles in the modulation of a wide range of cellular processes like proliferation, apoptosis, cell migration, cell polarization, membrane trafficking, cytoskeleton rearrangements, and transcriptional regulation. Different Rho GTPases are known to participate in normal animal development. However, it is also known that the deregulation of their activities contributes to the generation of different human pathologies, including cancer progression, neurodegenerative disorders, and infectious diseases. Although the Rho GTPase family is reported to contain at least 20 members in the human genome, research has focused mainly on 3 members, *CDC42, RAC1,* and *RHOA* [3,4]. For the rest of the proteins little or no functional information is currently available [3]. To gain further understanding of the biological function of Rho GTPases, the use of a suitable animal model such as zebrafish will be of great importance.

In this paper we report the presence of at least 23 Rho GTPases in the human genome. We also show that at least 32 Rho genes exist in zebrafish, representing one or more

☆ Supplementary data for this article may be found on ScienceDirect.
☆☆ Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. AY865555–AY865573 and AY965253.

* Corresponding author. Institute of Biology, Leiden University, Wassenaarseweg 64, 2333 AL Leiden, The Netherlands. Fax: +31 71 5274999.
E-mail address: salas@rulbim.leidenuniv.nl (E. Salas-Vidal).

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.03.010

counterparts for 17 human genes. We show that most of these genes are expressed in wild-type healthy adults, but only a subset is expressed early in zebrafish development. This suggests that these genes perform critical functions during the first steps of zebrafish development. In adult fish, the expressed genes might perform a more diverse set of functions. A role in disease is suggested for one of the genes, *rhogb,* which was found to be upregulated during *Mycobacterium marinum* infection.

Results

RHO GTPases in the human genome

To identify the Rho GTPases in zebrafish, we first compiled and analyzed the information available for the Rho GTPases in the main human genomic databases, NCBI and Ensembl. We used information from human genes that are located at different loci and that are supported by mRNA and EST evidence. According to these criteria 26 Rho GTPases genes exist in the human genome (Supplementary Table 1); however, for three novel sequences, two *RAC1* homologs (*RAC1P2* and *RAC1P4*) and one *RHOQ* homolog (*LOC284988*), only predicted mRNAs are available. Theoretically they encode full-length sequences, yet we could not find distinctive EST evidence for their transcription. In fact, the two *RAC1* homologs have been recently reported to be pseudogenes [5]. In the case of *LOC284988* the lack of reported ESTs also suggests that this is a pseudogene. Therefore, in the present work we used 23 human Rho GTPases genes in the following analyses.

Most human Rho GTPase proteins are around 200 amino acids (aa) in size; however, six members are larger, ranging from 400 to 700 aa. The three RhoBTBs and the two RhoTs showed additional C-terminal domains, such as the BTB domain in the case of RhoBTBs or the second GTPase domain and EFh domain characteristic of RhoT genes [6,7]. We used the NCBI Conserved Domain Search engine to verify that all 23 proteins are Rho members, since the insert region in the GTPase domain characteristic of Rho proteins is a divergent feature for some members, like RHOBTB3 and the two RHOT proteins [3]. This analysis gave in all cases as first and most statistically significant hit the Rho domain. Other additional computational tools were used to identify sequence signatures indicative of potential subcellular protein localization and known lipid modifications (Supplementary Table 2). Interestingly the RhoT proteins present hydrophobic stretches that suggest a possible transmembrane localization domain (Supplementary Table 2). In the yeast homolog to RHOT1 a similar carboxy-terminal domain has been experimentally shown to target these proteins to the mitochondrial outer membrane [8]. Other important domains present in RHOU and RHOV are the proline-rich regions proposed to serve as Src-homology 3 domain recognition sequences [9]. These are located at the amino-terminal sequence before the GTPase domain (Supplementary Table 2).

Phylogenetic analysis by the neighbor-joining method with bootstrapping (Supplementary Fig. 1) and maximum-likelihood analysis (not shown) produced consistent phylogenetic relationships among the different human Rho GTPases. Furthermore, the intron–exon structure of human Rho GTPases confirmed these relationships, since the structure was strikingly conserved in the most closely related genes (Fig. 2). For example, the number of introns and exons and the sizes of exons were highly preserved in the following cases: *RHOA* compared to *RHOC; RHOD* compared to *RHOF, RND1, RND3,* and *ARHN; RAC1* compared to *RAC2* and *RAC3; RHOJ* compared to *RHOQ; RHOU* compared to *RHOV; RHOBTB1* compared to *RHOBTB2; RHOT1* compared to *RHOT2* (Fig. 2). In general the central coding exons are the most conserved and the 5′ and 3′ coding exons have a more variable size and correlate with the hypervariable amino- and carboxy-terminal regions of the coded proteins. Retrotranscription followed by insertion into the genome seems to have occurred in at least two members of the Rho GTPases. *RHOB* is a single-exon-coded functional gene, which seems to have arisen more recently than *RHOA* and *RHOC,* by retrotranscription and insertion. In the case of the *RAC* genes, the two possible pseudogenes *RAC1P4* and *RAC1P2* and two additional genes described by Kluger and collaborators in 2004 [5] also seem to have arisen by the same mechanism.

Based on the phylogenetic analyses and the gene structure, and taking into account previous classifications supported by functional information [3], we are able to group the Rho GTPases into seven subfamilies: Rho, Rac, Cdc42, RhoD, Rnd, RhoBTB, and RhoT. One GTPase, RHOH, was not included in these subfamilies (Fig. 3).

Prediction of Rho GTPase genes in the zebrafish genome

The protein reference sequences of the 23 human Rho GTPases were used as queries for BLAST searches in the translated genomic zebrafish DNA. This search resulted in 65 zebrafish contigs with regions of homology to the human genes (as cutoff threshold we used 10^{-10}). The GENSCAN predictions were subsequently used as queries for BLAST searches against the whole set of human proteins available in the NCBI database. This analysis showed that only 36 represented unique putative Rho GTPase sequences, while others represented duplications or members of the Ras family of GTPases. Four putative Rho GTPase sequences were discarded because the predicted proteins contained sequences of unrelated genes, which could be the result of misplaced sequences. The remaining 32 predicted zebrafish Rho GTPases showed a high percentage identity with specific members of the human Rho GTPase family (more than 65% identity, except for RHOBTB3 with 40%) (Table 1). The primary amino acid sequences of these zebrafish

proteins were subjected to the same sequence analysis methods as used for the human homologs. Based on the identification of sequence motifs by the PSORT II software [10], such as consensus cleavage site motifs in mitochondrial targeting peptides, endoplasmic reticulum, peroxisomal localization signals, and C-terminus prenylation modifications, predictions were made (Supplementary Table 3 and Table 1). To analyze in more detail the conserved regions between the zebrafish and the human genes we performed a multiple amino acid sequence alignment of the GTPase domain (Fig. 1) in addition to global alignments of the complete sequences (not shown). These global and local alignments were further used to perform phylogenetic analysis by two alternative methods, neighbor-joining (Fig. 3) and maximum-likelihood (not shown), which gave consistent results. All the above information was integrated to assign the identity of each zebrafish Rho GTPase. Relevant details of the analyses of the different subfamilies are presented in the following sections. The names assigned to the zebrafish Rho GTPases follow the nomenclature guidelines of ZFIN (http://zfin.org).

Rho subfamily

Five contigs encoding predicted putative full-length Rho subfamily genes were identified (Table 1). The overall protein sequence identity of all human and zebrafish Rho subfamily members is 79.1% and the overall similarity is 97.4%. BLAST searches of the predicted proteins gave the most significant identity and *E* values for human RHOA and RHOC. However, detailed analysis of the intron–exon organization and splicing sites showed that the five predicted genes are similar to *RHOA,* with one 5′ noncoding exon followed by four coding exons (Fig. 2 and Supplementary Fig. 1a). The conservation of the central coding exons is outstanding up to the number of base pairs and the intron phase, which is the position of the intron within a codon. Based on these analyses, zebrafish contains five homologs of *RHOA*. In accordance, these genes were named *rhoaa, rhoab, rhoac, rhoad,* and *rhoae* (Table 1).

Rac subfamily

Seven contigs encoding putative full-length Rac subfamily genes were identified (Table 1). The overall protein sequence identity of all human and zebrafish RAC family members is 45.8% and the overall similarity is 98.4%. In BLAST searches four of the predicted proteins gave the most significant identity and *E* values for human RAC1, RAC2, and RAC3. In three cases, the predicted exon organization also was similar to that of *RAC1, RAC2,* and *RAC3,* which each have six coding exons. One zebrafish gene, similar to *RAC1,* contained only a single coding exon (Fig. 2 and Supplementary Fig. 1b). Based on the BLAST and the two phylogenetic analysis methods used, zebrafish contains two counterparts of RAC1 and single counterparts of RAC2 and RAC3 (Table 1 and Figs. 2 and 3). In accordance, these genes were named *rac1a, rac1b, rac2,* and *rac3* (Table 1). Like the human proteins, all full-length RAC homologs contained nuclear localization signals and potential prenylation modification signals.

Three zebrafish predicted Rac subfamily proteins gave human RHOG as first hit in BLAST searches and displayed a gene structure similar to that of the human gene (Fig. 2 and Supplementary Fig. 1b). However, the zebrafish RHOG homologs diverge from the single human RHOG protein in that they lack the nuclear localization signal (Supplementary Table 3). Because the classification of these sequences was supported by phylogenetic analysis and gene structure (Figs. 2 and 3), we assigned the names *rhoga, rhogb,* and *rhogc*.

Cdc42 subfamily

Seven contigs encoding putative Cdc42 subfamily genes were identified. Five contigs comprise complete coding sequences and two comprise only partial sequences. The overall identity of all human and zebrafish family members is 20.6% and the overall similarity is 63.7%. Based on BLAST searches against human proteins, phylogenetic analysis, and largely conserved exon organization (Table 1, Fig. 2, and Supplementary Fig. 1c) we could assign three sequences as CDC42 homologs, one as a RHOJ homolog, two as RHOU homologs, and one as a RHOV homolog. No RHOQ homolog has been found.

One of the three CDC42 homologs was 99% identical to human CDC42 and contained nuclear localization and prenylation modification signals. We assigned to this gene the name *cdc42a*. The other two predicted proteins, named Cdc42b and Cdc42c, were 77 and 90% identical to human CDC42, respectively. The Cdc42b predicted protein does not contain a recognizable nuclear localization signal and is shorter than the human homolog. The Cdc42c predicted protein showed multiple nuclear localization signals compared to the human protein.

One zebrafish gene sequence was assigned as *rhoj* and the predicted protein contained conserved nuclear localization and prenylation modification signals compared to human RHOJ. The zebrafish RHOU homologs, named Rhoua and Rhoub, showed the proline-rich domains at the N-terminus; however, they lack the C-terminal prenylation modifications signals compared to the human homolog (Supplementary Tables 2 and 3). The zebrafish RHOV homolog was named Rhov and an interesting difference is that it does not have the proline-rich domain (Supplementary Table 3).

RhoD subfamily

One contig encoding a putative RhoD subfamily gene was identified. The predicted sequence available is partial at the amino-terminal region. Based on BLAST search, phylogenetic analysis, and conserved exon organization (Table 1, Fig. 2, and Supplementary Fig. 1d), the sequence found is assigned as a RHOF homolog. The overall protein identity of the human and zebrafish members in this subfamily is 42.9% and the overall similarity is 84.9%.

Table 1
Zebrafish Rho small GTPases analysis

Protein name in *Danio rerio*	Alternative common names	Contig[a]	Sequence accession numbers	Predicted protein size[b]	Rho domain position and other domains[c]	CaaX/ modification[d]	BLAST against human proteins[e]	Identity positives[e]	Phylogenetic analysis[f]
Rhoaa	RhoA, ARH12, ARHA, RHO12, RHOH12	BX004884	NM_213137.1 AY865555[g]	193	3–181	CALL/G	RHOA e–103	93–96%	RHOA
Rhoab		BX784025	NM_212749 AY865556[g]	193	3–181	CCLL/G	RHOA e–105	95–96%	RHOA
Rhoac		Zv4_scaffold1409.3	NM_213350	193	3–181	CLLL/G	RHOA e–106	96–97%	RHOA
Rhoad		BX248319	NM_001002445 AY865557[g]	193	3–180	CLLL/G	RHOA e–105	94–97%	RHOC/ RHOA
Rhoae		Zv4_NA8994.1	NM_201150 AY865558[g]	193	3–180	CSLL/G	RHOC/ RHOA e–102/ e–102	92–95% 92–95%	RHOC/ RHOA
Rnd3a	RhoE, Rho8, Rnd3, RHOE	Zv4_scaffold212	NM_199522 AY865559[g]	243	21–197	CTVM/F	RND3 e–121	90–95%	RND3
Rnd3b		CR394542	NM_001002591 AY865560[g]	243	9–183	CTVM/F	RND3 e–95	76–87%	RND3
Rhof	RhoF, Rif, FLJ20247, ARHF	BX088560 BX005309	CD760128 Predicted	260	17–233	CTVL/G	RHOF e–71	68–82%	RHOF
Rhoga	RhoG, TVHURG, ARHG	BX000526.8	NM_200680.1 AY865561[g]	191	6–164	CVLL/G	RHOG e–86	77–87%	RHOG
Rhogb		CR391971	NM_200040 AY865562[g]	191	6–179	CILL/G	RHOG e–86	77–87%	RHOG
Rhogc		BX005407.5	NM_199692 AY965253[g]	191	1–179	CVLL/G	RHOG e–87	77–86%	RHOG
Rhoh	RhoH, TTF, ARHH	AL844559	Prediction AY865563[g]	188 (p, lacking N)	1–163	NF	RHOH e–74	79–90%	RHOH
Rhoj	ARHJ, RhoJ, TC10-like, TCL, TC10B, FLJ14445	BX248118	AL915698.1	197 (p, lacking N)	2–179	CALV/G	RHOJ e–100	91–95%	RHOJ
Rhoua	WRCH1, hG28K, WRCH-1, CDC42L1, FLJ10616, ARHU	AL772388	NM_001007443 AY865564[g]	253	44–221	NF	RHOU e–94	81–91%	RHOU
Rhoub		Zv4_scaffold1202	AY865565[g]	235	29–203	NF	RHOU e–77	71–84%	RHOU
Rhov	Chp, WRCH2, RHOV	BX897725.7	CR759734.3	209 (p, lacking N)	4–167	NF	RHOV e–73	64–77%	RHOV
Cdc42a	G25K, CDC42Hs	Zv4_scaffold305	CK026369.1 AY865566[g]	191	1–179	CCIF/G	CDC42TV2 e–104	99–99%	CDC42
Cdc42b		AL929578.4	BM316198.1	185	1–173	CVIT/G	CDC42TV1 e–76	767–86%	CDC42
Cdc42c		BX511250	NM_199865.1 AY865567[g]	191	1–178	CVLL/G	CDC42TV2 e–99	90–97%	CDC42
Rac1a	TC-25, p21-Rac1	Zv4_scaffold1035	NM_199771 AY865568[g]	192	1–175	CLLL/G	RAC1 e–101	100–100%	RAC1
Rac1b		BX537286.4	CK026369.1	192	1–179	CLIL/G	RAC1 e–75	70–80%	RAC1
Rac2	Gx, EN-7, HSPC022	BX571960.6	NM_001002061 AY865569[g]	192	4–182	CVML/G	RAC2 e–104	93–98%	RAC2

Table 1 (*continued*)

Protein name in *Danio rerio*	Alternative common names	Contig[a]	Sequence accession numbers	Predicted protein size[b]	Rho domain position and other domains[c]	CaaX/ modification[d]	BLAST against human proteins[e]	Identity positives[e]	Phylogenetic analysis[f]
Rac3		BX897685	BC076433 AY865570[g]	192	1–179	CTVF/G	RAC3 e–109	97–99%	RAC3
Rnd1a	Rnd1, RHO6	Zv4_scaffold1916.1	BC076165 AY865571[g]	233	12–190	CSVM/F	RND1 e–91	82–90%	RND1
Rnd1b		BX855597.6	NM_212854 AY865572[g]	231	13–190	CTIM/F	RND1 e–100	78–90%	RND1
Rhobtb1a	KIAA0740, MGC33059, MGC33841	Zv4_NA13247.1 Zv4_scaffold402.9	Prediction	593 (p)	2–110 BTB 265–352/ 372–475	NF	RHOBTB1 e0	68–80%	hRHOBTB1
Rhobtb1b		BX784028	AL923174	267	15–207 No BTB domain found	NF	RHOBTB1 e–129	83–89%	RHOBTB1
Rhobtb2a	DBC2, KIAA0717	Zv4_scaffold810.1	CO354542 AY865573[g]	684	13–207 BTB (2 hemidomains) 383–443/ 494–555 2nd 568–603	NF	RHOBTB2 e0	76–83%	RHOBTB2
Rhobtb2b		Zv4_scaffold1621.1	BI476282	294	14–206 BTB 264–443 2nd 464–566	NF	RHOBTB2 e–154	91–95%	RHOBTB2
Rhobtb3	KIAA0878	Zv4_scaffold1733	Prediction CO934269	591	Ras 59–183 BTB 387–497	CSIM/F	RHOBTB3 e–122	40–57%	RHOBTB3
Rhot1a	MIRO-1, FLJ11040, ARHT1	AL954746.8	Prediction BI427999	632	RHO 6–170 Rab 491–554 EFh 189–217/ 309–337 Transmembrane 571–593/ 606–628	NF	RHOT1 e0	80–90%	RHOT1
Rhot1b		BX663498.3	Prediction	291(p)	Rab 200–263 EFh 18–46	NF	RHOT1 e–138	75–87%	NI

[a] Zebrafish genomic contig in the Ensembl database comprising the coding sequence of the predicted protein.
[b] Prediction by GENSCAN using the genomic sequence and/or based on the predictions contained in Ensembl.
[c] Prediction according to NCBI Conserved Domain Search and SMART.
[d] Prediction from PSORT II analysis. NF, not found.
[e] BLASTP at NCBI.
[f] Neighbor-joining and maximum-likelihood phylogenetic analysis performed with ClustalX and TreePuzzle.
[g] Sequences generated in the present article.

The human RHOF and zebrafish Rhof proteins showed similar nuclear localization and C-terminal prenylation modification signals. Human RHOF has been reported to show primary sequence differences compared to CDC42 that should affect the interaction with the CRIB (CDC42-RAC interactive binding) domain of downstream effectors. In particular, human RHOF shows divergence in two regions, called switch 1 and α5 helix, compared to CDC42. In CDC42 the amino acids important for CRIB domain interaction are threonine 35, aspartic acid 38, valine 42, and leucine 174. In RHOF these positions are substituted with serine, glutamic acid, alanine, and lysine. The substitutions in human RHOF and zebrafish Rhof make it unlikely that these interact with CRIB domain proteins [11]. In the zebrafish Rhof three of these four positions are changed compared to CDC42, except for the aspartic acid 38, which is conserved (Supplementary Fig. 1d, asterisks).

Rnd subfamily

Four contigs encoding putative full-length proteins with significant homology to human Rnd subfamily genes were identified (Table 1). The zebrafish proteins represented two RND1 homologs and two RND3 homologs, named Rnd1a and Rnd1b and Rnd3a and Rnd3b. This classification was supported by BLAST analysis, phylogenetic analysis, and predicted exon organization of the genes, except that one RND1 homolog contained an extra 3′ noncoding exon (Fig. 2). The two zebrafish *rnd3* genes lacked the 5′ noncoding exon present in the human

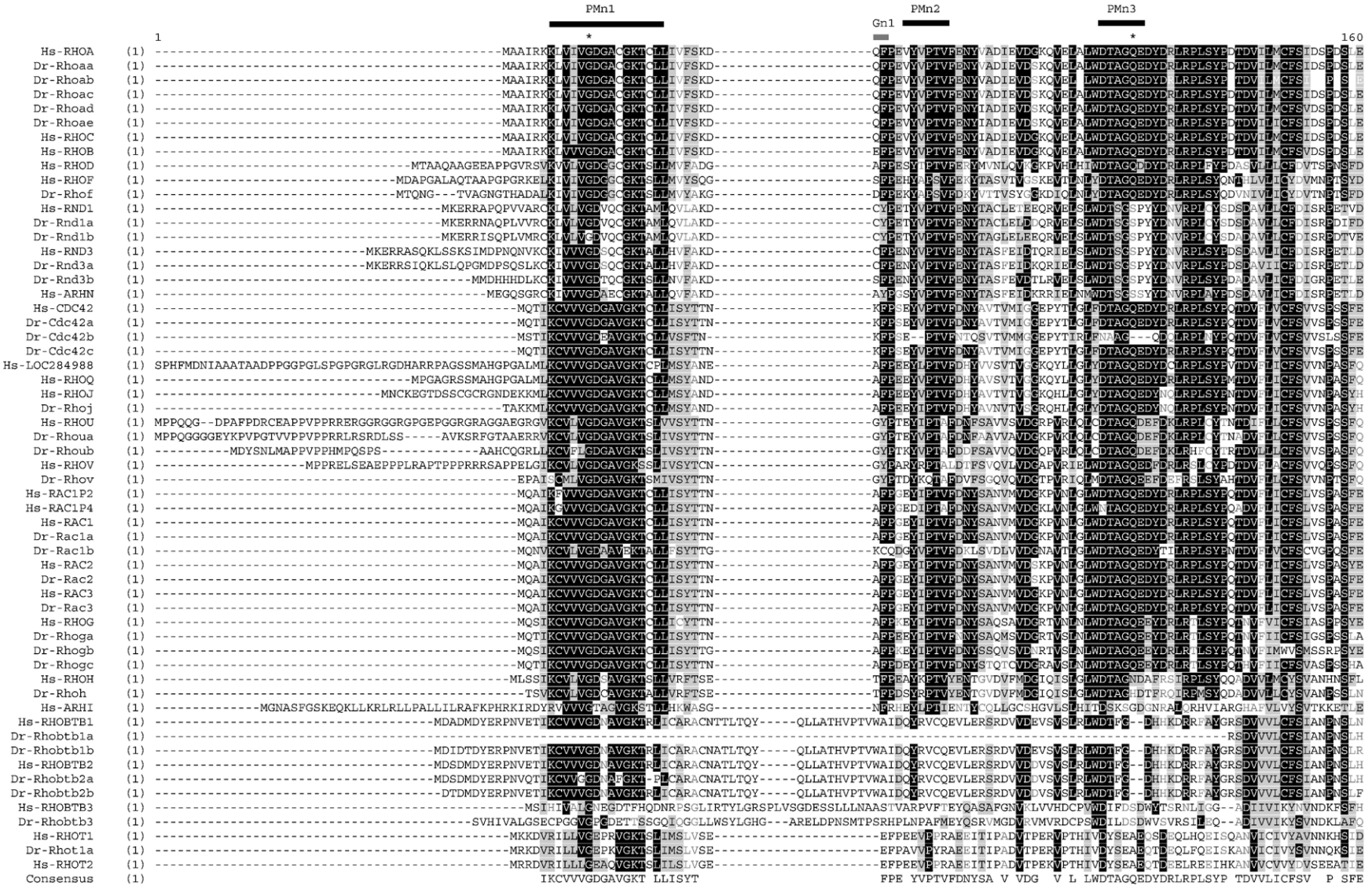

Fig. 1. Amino acid alignment of the Rho GTPase domains of human and zebrafish proteins. Black horizontal lines, consecutive phosphate/magnesium binding regions PMn 1, PMn 2, and PMn 3. Gray horizontal lines, consecutive guanine base binding regions Gn1, Gn2, and Gn3. Asterisks, amino acids essential for GTP hydrolysis. Gray dotted line, Rho insert region. Black dotted line, CaaX box. Block of similarity, black text in gray background. Conservative changes, white text in black background. Weak similarity changes, gray text.

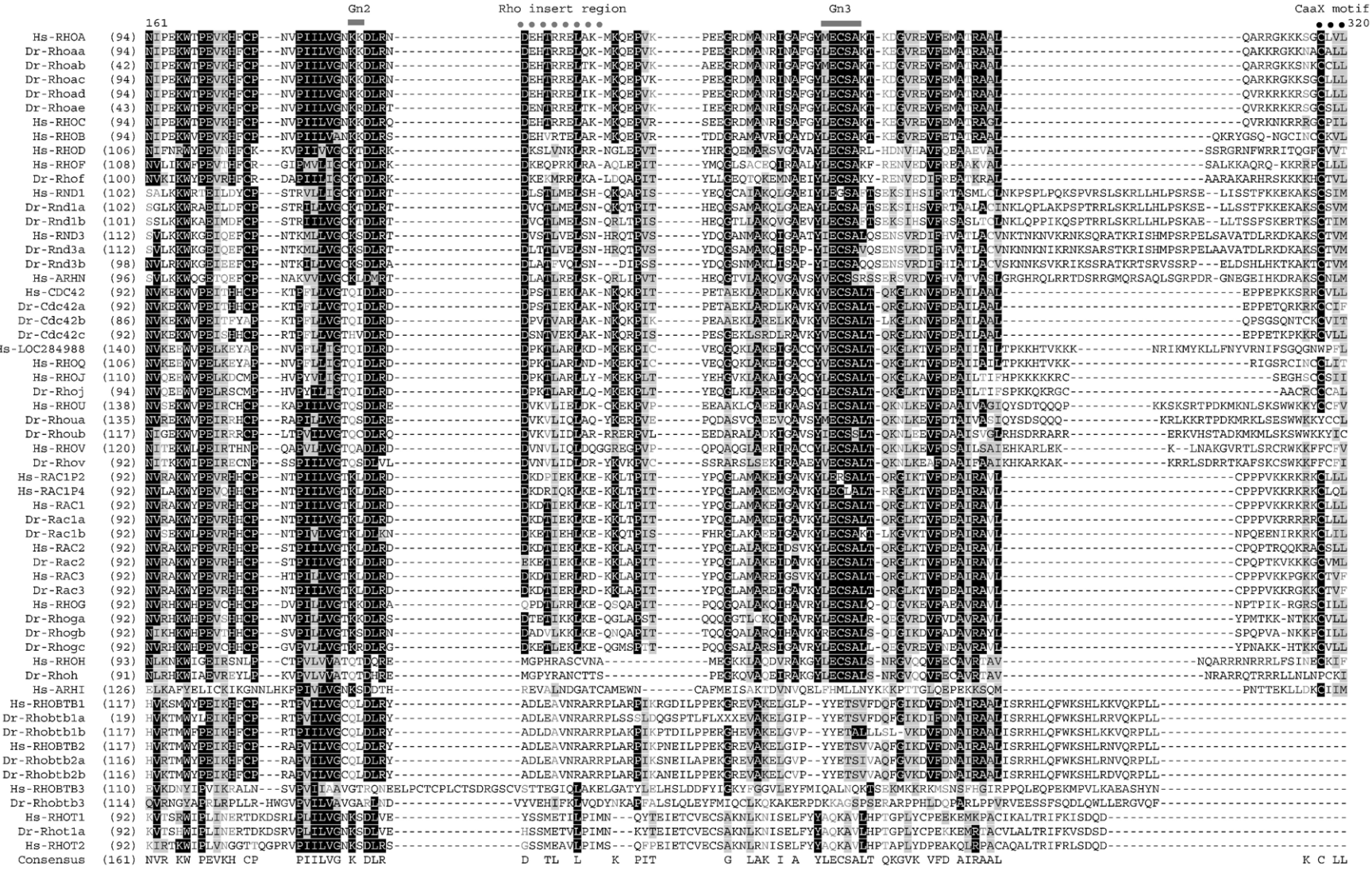

Fig. 1 (*continued*).

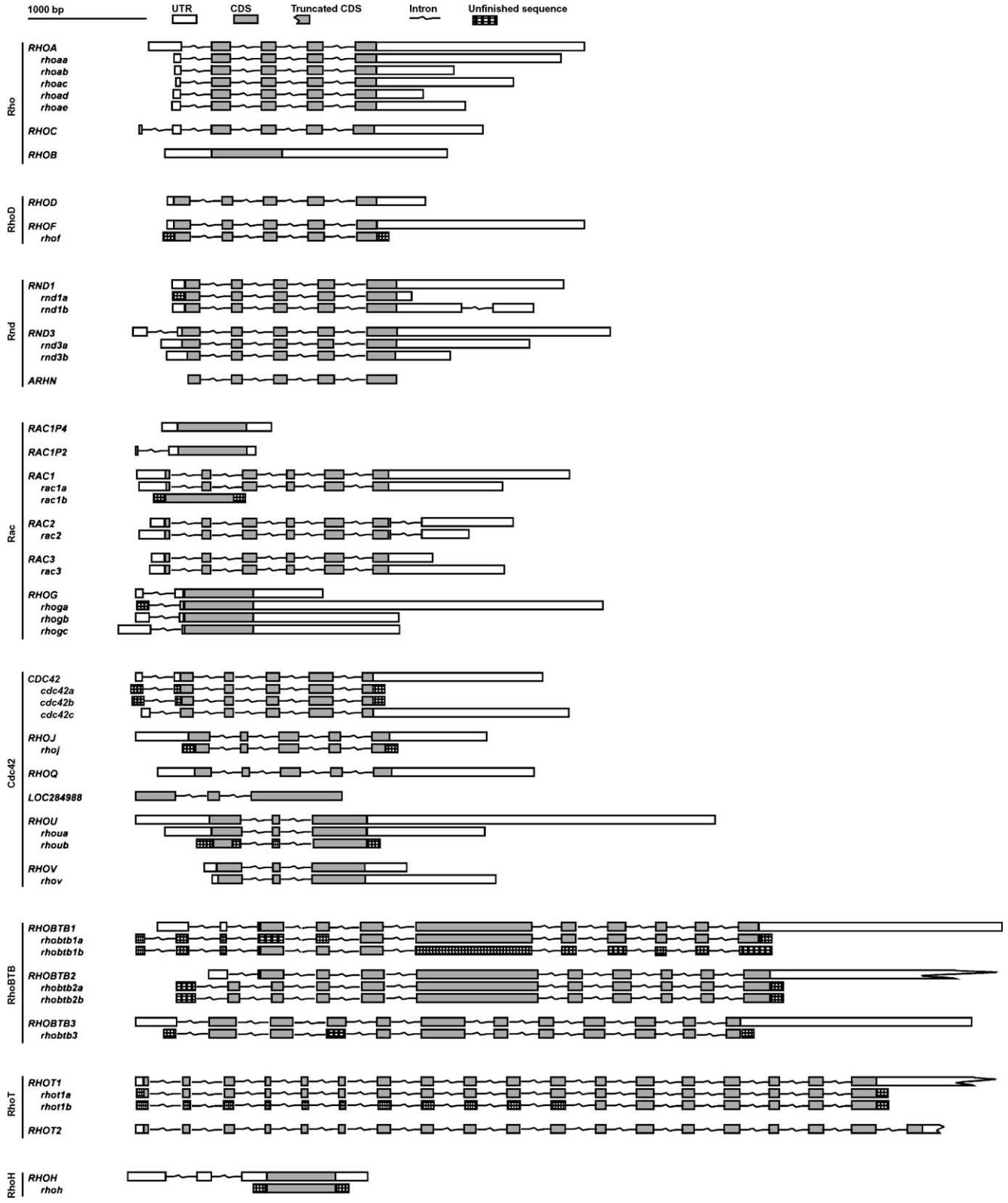

Fig. 2. Schematic representation of intron–exon organization of human and zebrafish Rho small GTPase genes. Shown in gray are the coding regions. The introns are collapsed for presentation purposes.

homolog (Fig. 2). The overall identity of all human and zebrafish protein family members is 36.1% and the overall similarity is 88.9%.

Human Rnd subfamily genes show several distinctive features that were also found in the zebrafish counterparts. Zebrafish predicted proteins contain C-terminus prenylation

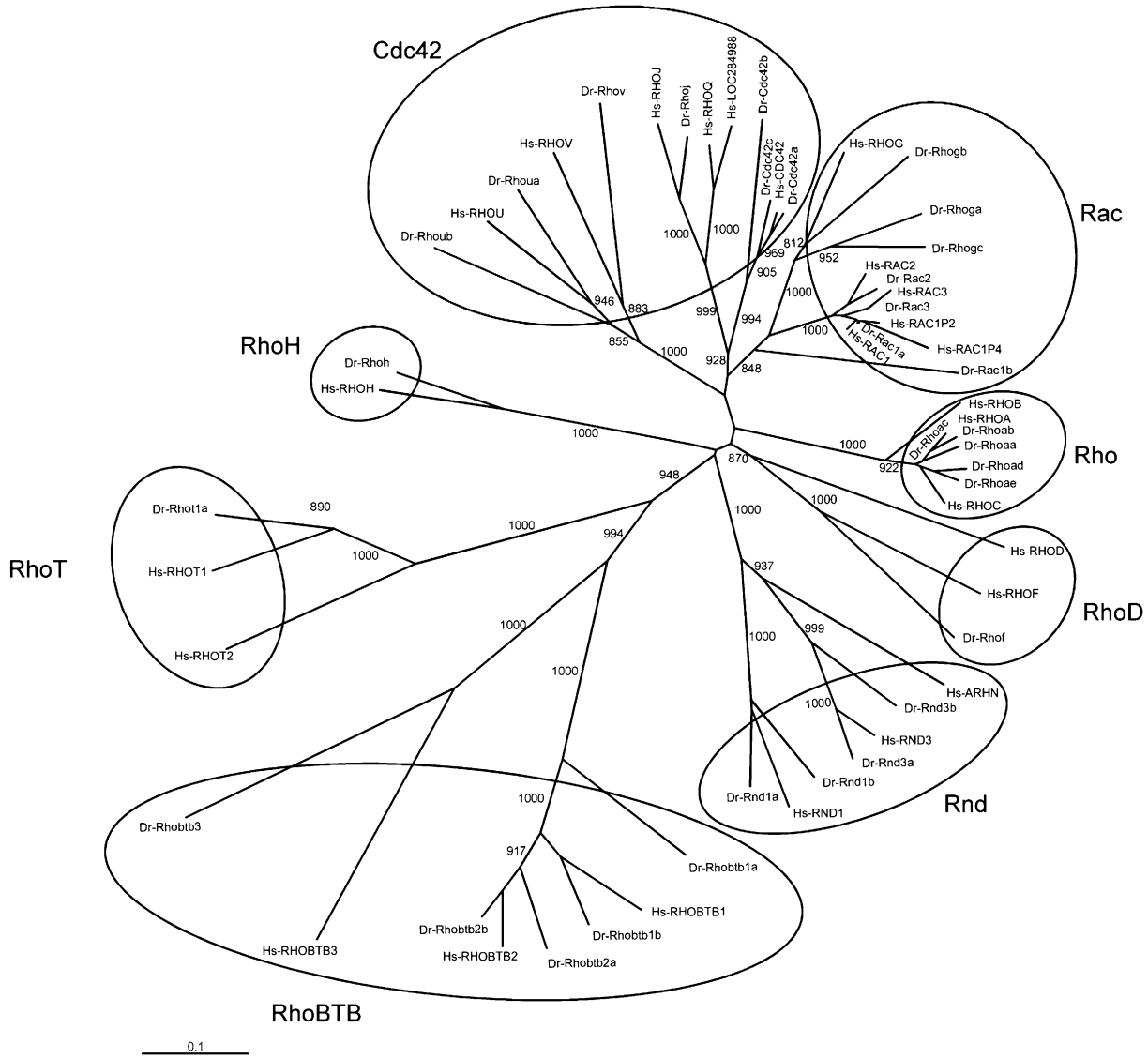

Fig. 3. Unrooted phylogenetic tree of human and zebrafish Rho GTPases. Human proteins are in capitals; zebrafish protein names are in lowercase. The tree was constructed by neighbor-joining analysis based on an alignment of the amino acid sequences of the Rho domains. The numbers indicate the occurrence of nodes during bootstrap analysis with 1000 reiterations. Only values above 800 are shown.

motifs that end with methionine, like the human homologs, and therefore are likely to be farnesylated as well. When the RND genes were originally described it was also found that these proteins show substitutions compared to the RAS genes that lead RND proteins to be constitutively active proteins by decreasing their intrinsic GTPase activity and preventing GAP-mediated GTPase stimulation [12]. Compared to RAS the glycine 12 is replaced by a valine in both predicted zebrafish RND1 homologs as is found in the human RND1 protein. This position in human RND3 is substituted by a serine, as in one of the zebrafish homologs, and by a threonine in the other. The glycine 13 found in RAS is replaced by glutamine in all zebrafish RND and RND3 homologs. Finally, the RAS alanine 59 and glutamine 61 are both replaced by serine in all predicted zebrafish proteins, like in the human homologs (Supplementary Fig. 1e, asterisks).

RhoH subfamily

BLAST searches in the zebrafish genome revealed at least one contig that encodes a predicted protein with significant homology with the human RhoH subfamily gene (Table 1). The predicted sequence is partial at the amino-terminal region. BLAST searches of the zebrafish predicted protein against human proteins gave the most significant identity and *E* values for RHOH, which was supported by phylogenetic analysis. Also we found that the predicted exon organization was similar to that of human *RHOH,* that is, with two noncoding exons and one coding exon (Fig. 2 and Supplementary Fig. 1f). Therefore this sequence was named *rhoh*. The overall identity and similarity of zebrafish protein to human RHOH is very high (72.8 and 98.4%, respectively).

Interestingly, RHOH has been reported to show particular amino acid substitutions compared to RAS

[13]. In RHOH the normally conserved residue glycine 12 is substituted for a serine in the human protein and by a cysteine in the zebrafish protein. Also the conserved glutamine 61 found in RAS is substituted by an asparagine in the human RHOH and substituted by a histidine in zebrafish Rhoh (Supplementary Fig. 3f, asterisks).

RhoBTB subfamily

Five contigs containing partial coding sequences with significant homology to human RhoBTB subfamily genes were identified (Table 1). BLAST searches and phylogeny reconstruction identified two potential RHOBTB1 homologs, which we named Rhobtb1a and Rhobtb1b; two RHOBTB2 homologs, named Rhobtb2a and Rhobtbt2b; and one RHOBTB3 homolog, named Rhobtb3. The intron–exon organization is highly conserved between human and zebrafish although some inconsistencies were found, mainly in predictions from unfinished contigs (Fig. 2 and Supplementary Fig. 1g). When the Rho GTPase domain is compared between the RHOBTB1 and RHBTB2 proteins from human and the Rhobtb1b, Rhobtb2a, and Rhobtb2b of zebrafish, the identity is 77.4% and the similarity 99%. RHOBTB3 from human and zebrafish is more divergent, showing 39.8% overall identity and 54.2% similarity.

All human RHOBTBs show two BTB domains in their primary protein sequence (see Supplementary Table 1). In the case of the zebrafish proteins it is more variable, at least based on the partial sequence information available. Rhobtb1a, Rhobtb2a, and Rhobtb2b show the two BTB domains, while Rhobtb1b does not show any BTB domain and might represent a truncated protein. Rhobtbt3 showed only one predicted BTB domain (Table 1). Primary sequence analysis of human proteins showed that all tree RHOBTBs have nuclear localization sequences and no other potential modifications, except human RHOBTB3, which contains a C-terminal prenylation motif that ends with methionine and therefore is likely to be farnesylated. Only zebrafish Rhobtb2a showed a nuclear localization signal, and the predicted zebrafish Rhobtbt3 showed a prenylation motif ending with a methionine as the human homolog (Table 1 and Supplementary Fig. 3g).

RhoT subfamily

Two contigs encode predicted proteins with significant homology with human RhoT subfamily genes (Table 1). One of the predicted sequences appears to be complete and one seems to be partial. BLAST searches and phylogenetic analysis identified both sequences as RHOT1 homologs and they were consequently named Rhot1a and Rhot1b. No RHOT2 was found. As is the case for the RHOBTBs these sequences were of poor quality but the intron–exon organization seems to be conserved. When only the Rho GTPase domains are compared we found that the identity is 71.5% and the similarity is 98.2%.

The two predicted zebrafish RHOT proteins showed the characteristic second GTPase domain and the expected EFh domain as found in the human homologs. It is important to mention that Rhot1b was removed from the alignment shown in Fig. 1 and the phylogenetic tree because it is a partial sequence and contained only the second GTPase domain and not the first, which is expected to be present within all the Rho GTPases. Another distinguishing feature from the RHOT1 gene is that the conserved glycine 12 and glutamine 61, found in other GTPases like RAC1, RHOA, and CDC42, are substituted by a glutamic acid and by alanine, respectively [7]. In the zebrafish Rhot1 protein for which we have information about the first GTPase domain, this position shows the same substitutions (Supplementary Fig. 1h, asterisks).

Expression analysis of zebrafish Rho GTPase genes

We analyzed whether the predicted Rho GTPase genes are expressed during zebrafish development by RT-PCR. Where possible we designed oligonucleotides in different exons (Supplementary Table 4). In addition, we treated the RNA samples with DNase to avoid genomic amplification instead of cDNA amplification. By RT-PCR analysis using total RNA extracted from adult fish we could demonstrate the expression of 20 of the predicted genes (Fig. 4A). The identity of the amplified RT-PCR products was confirmed by cloning and sequencing of the products. We were not able to amplify 12 of the predicted genes by RT-PCR (*rhoac, cdc42b, rhof, rhoj, rac1b, rhov, rhobtb1a, rhobtb1b, rhobtb2b, rhobtb3, rhot1a,* and *rhot1b*); however, searches

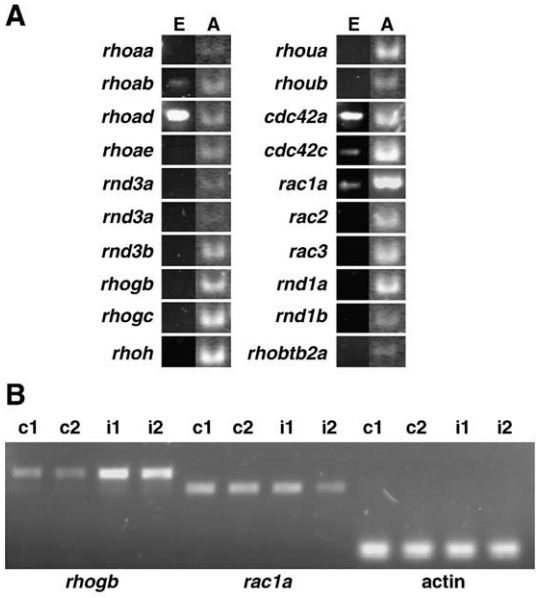

Fig. 4. RT-PCR analysis of zebrafish Rho gene expression. (A) Expression analysis in embryos at 30% epiboly and in adult zebrafish. (B) Expression analysis after *M. marinum* infection. RNAs used for amplification were from adult male zebrafish infected (i1 and i2) by intraperitoneal inoculation with *M. marinum* strain M (fish i1) or strain E11 (fish i2) or from healthy fish inoculated with control fluid (c1 and c2). β-Actin was used as a control for constitutive expression.

Table 2
Zebrafish *rhogb* is upregulated in response to *M. marinum* infection

	Affymetrix	Spotted oligonucleotide set
UniGene cluster	Dr.9665	Dr.9665
GenBank accession No.	BC044508	BG307536
Fold change	5.2	4.0
p value	0	0.00015

Two microarray types, Affymetrix GeneChips and spotted oligonucleotide sets (designed by Sigma–Compugen) from previously published results from our laboratory [15], were used to confirm the upregulation of zebrafish *rhogb*.

for reported ESTs in the NCBI server show ESTs for all of these genes except for *rhobtb1a* and *rhot1b*.

Interestingly, when we performed RT-PCR from RNA extracted from zebrafish embryos at 30% epiboly, which is just prior to the onset of gastrulation, only five genes were found to be expressed, *rhoab, rhoad, cdc42a, cdc42c,* and *rac1a* (Fig. 4A).

Expression analysis in Mycobacterium-infected zebrafish

To analyze if zebrafish Rho small GTPase genes are responsive to mycobacterial infection we compared the expression levels of *M. marinum*-infected fish against those of healthy control fish. RNA was isolated from fish after 8 weeks of intraperitoneal injection of mycobacteria suspension. These fish showed distinctive tuberculosis disease symptoms [14]. Interestingly, from all the Rho small GTPases tested, only *rhogb* showed a reproducible upregulation of expression after tuberculosis infection compared with control fish (Fig. 4B). Similar expression levels were observed with the other Rho small GTPases tested and the reference gene β-actin. For comparison *rac1a* expression, which does not change after infection with *Mycobacterium,* is shown. We have confirmed the upregulation of *rhogb* by analysis of our published microarray results obtained using two independent technologies: Affymetrix gene chips and spotted oligonucleotide microarrays [15] (Table 2).

Discussion

We have made an extensive analysis of the information available for the human genome and the Sanger zebrafish sequencing project to identify and annotate the homologous Rho small GTPases. We found that the zebrafish genome contains at least 32 Rho genes representing one or more homologs of 17 of the 23 predicted genes in human. In a phylogenetic analysis all zebrafish genes clustered within the different human subfamilies. Furthermore we confirmed the expression of 20 zebrafish genes using RT-PCR analysis. For another 10 genes we found corresponding partial EST sequences in public databases such as NCBI. For 2 genes we did not find any evidence for their transcription and these therefore might represent pseudogenes.

The Sanger Zv4 genome release used in this study covers the whole genome and about two-thirds of it is based on finished genomic sequences. Since we used as queries all the human genes to search the zebrafish genome, it is likely that we have analyzed most of the members of this gene family present in zebrafish. Furthermore, the predictions of the majority of the Rho GTPases reported in this paper are based on the finished Zv4 genome sequence information. About one-third of the gene predictions are based on genome sequences of unfinished regions. Some of these predictions represent partial sequences and it cannot be excluded that some may contain misplaced sequence parts. However, we reported these sequences because we found evidence for their expression. We estimate that the Rho GTPase genes reported in this study represent at least 90% of all Rho GTPases present in the zebrafish genome.

The identification of zebrafish homologs for most human Rho GTPases subfamilies indicates that these Rho subfamilies should have been present in the ancestral precursor and probably confer selective advantages since fish separated from other vertebrates approximately 430 million years ago [16]. The most conserved gene subfamilies between zebrafish and human are Rho, Rac, and Cdc42 (Figs. 2 and 3). In particular some members show an outstanding degree of identity in the protein sequences among homologous human counterparts, like Rhoaa, Rhoab, Rac1a, Cdc42a, and Cdc42c, with identity percentages higher than 90% and similarities reaching almost 100%. Only two other proteins from two other subfamilies showed an identity percentage higher than 90%: Rnd3a and Rhoh. This degree of conservation indicates that their functions are important in all vertebrates. In addition, there is a remarkable intron–exon organization and intron phase conservation (not shown) among subfamily members, which suggest this is a highly evolutionarily constrained feature. Our analysis also suggests that retrotranscription and insertion events occurred in zebrafish in *rac1b* giving rise to a single exon-coded sequence.

In early zebrafish development only a small subset of GTPases are expressed, and these might represent the genes that play most critical roles before and around epiboly–gastrulation. The genes that were expressed are homologous to the best studied human Rho GTPases, *RHOA, CDC42* (two zebrafish homologs), and *RAC1*. *RAC1, CDC42,* and *RHOA* are known to be expressed in mouse early development with dynamic patterns of localization [17,18]. Knockout mice and morpholino knockdown studies in *Xenopus* showed that *RAC1* is required for the proper formation of the three germ layers during gastrulation [19,20]. Mouse knockout studies of *CDC42* also showed that embryos stop development even before gastrulation and degenerate at earlier stages than the *RAC1* knockout [21]. We have not been able to find knockout reports for *RHOA*. Currently we are characterizing the early developmental function of all the five small GTPases found to be expressed in early zebrafish embryos.

Rho GTPases are known to play important roles during the infection process of different pathogenic bacteria, including different *Mycobacterium* strains [22,23]. Previously we have characterized the human Toll-like receptor homolog genes in zebrafish that are important in the innate immune response to mycobacterial infections and found that their expression is upregulated in response to *M. marinum* infection [14]. It was striking to find that there was one Rho small GTPase, *rhogb,* that was affected in its expression during the course of *Mycobacterium* infection. The significant upregulation of this particular GTPase suggests that it plays a specific role in this type of bacterial infection. It has been previously reported that invasive *Salmonella enterica* serovar Typhimurium secrete effector proteins like SopE that participate in bacterial internalization. SopE is known to have a nucleotide exchange activity on different Rho small GTPases, including RHOG, and is able to activate these proteins [24]. Also RHOG is known to participate indirectly in the activation of other GTPases like RAC1 and CDC42, stimulating the formation of membrane ruffles and filopodia [25]. Interestingly, *RhoG* knockout mice show an in vivo modest increase in IgG levels and humoral response to antigen challenge and increase in T and B cell proliferation during in vitro stimulation of primary cell cultures, compared to wild-type mouse cells [26]. Therefore it is tempting to speculate that in zebrafish *rhogb* upregulation might affect Rac1 and Cdc42 activation, which might facilitate *Mycobacterium* infection and negatively modulate the immunological host response to the pathogen challenge. This hypothesis will be tested in our future research.

Materials and methods

In silico search for human and zebrafish putative Rho GTPases

We identified 23 different Rho GTPases in the human genome, and 3 potential pseudogenes, by searching the currently available main genomic databases at NCBI (http://www.ncbi.nih.gov/) and Ensembl (http://www.ensembl.org/). Only reference sequences supported by mRNA and EST evidence were selected, and it was shown that each corresponded to different genomic loci (Supplementary Table 1). The official gene names of the Rho GTPases were given according to the HUGO Gene Nomenclature Committee (http://www.gene.ucl.ac.uk/nomenclature/). These sequences were used as queries for searching for potential homologs in the zebrafish genome and EST collections, using the TBLASTN program at the Sanger Institute Ensembl BLAST server and NCBI.

The intron–exon boundaries of the human and zebrafish genes were analyzed by alignment of mRNA sequences with genomic sequences and by manual inspection of corresponding splicing sites. The protein domain structures were predicted using the NCBI Conserved Domain Search software (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and the SMART software (http://194.94.45.211) [27]. For subcellular localization and modification predictions we used PSORT II software (http://psort.nibb.ac.jp/form2.html) [10].

Sequence alignments and phylogenetic analysis

Global and local sequence alignments of the human and zebrafish Rho GTPase sequences were made using the program Vector NTI version 9.0 and manually adjusted. The dendrogram of zebrafish sequences and the phylogenetic tree were analyzed and constructed by two methods, the neighbor-joining method [28] using ClustalX 1.81 (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) [29] and maximum-likelihood analysis using TreePuzzle 5.0 software [30]. ClustalX analysis was done with default settings. Bootstrap sampling was reiterated 1000 times. For the matrix table "Gonnet" was used [31]. For pair-wise alignments the gap opening penalty was set to 35 and the gap extension penalty was set to 0.75. For multiple alignments the gap opening penalty was set to 15 and the gap extension penalty to 0.30, and divergent sequences alignment was delayed 30%. Trees were drawn using TreeView (http://taxonomy.zoology.gla.ac.uk/rod/treeview.html).

GenBank accession numbers of human sequences used for the alignments and phylogenetic tree analysis are shown in Supplementary Table 1.

RT-PCR and gene cloning

Total RNAs were isolated from zebrafish samples. RT-PCR experiments were performed and analyzed as described by Meijer and colleagues [14]. Sequences were deposited in the GenBank database under Accession Nos. *rhoaa* (AY865555), *rhoab* (AY865556), *rhoad* (AY865557), *rhoae* (AY865558), *rnd3a* (AY865559), *rnd3b* (AY865560), *rhoga* (AY865561), *rhogb* (AY865562), *rhogc* (AY965253), *rhoh* (AY865563), *rhoua* (AY865564), *rhoub* (AY865565), *cdc42a* (AY865566), *cdc42c* (AY865567), *rac1a* (AY865568), *rac2* (AY865569), *rac3* (AY865570), *rnd1a* (AY865571), *rnd1b* (AY865572), *rhobtb2a* (AY865573).

We used RNA samples from zebrafish infection experiments that were previously published [14].

Acknowledgments

This work was supported by a European Commission 6th Framework Programme grant (Contract LSHG-CT-2003-503496, ZF-Models). E.S.-V. was partially supported by a DGAPA/UNAM fellowship. The authors thank Astrid van der Sar and Wilbert Bitter (VU Medical Centre, Amsterdam) for providing *M. marinum*-infected zebrafish and Robert

Geisler (Max-Planck-Institut für Entwicklungsbiologie, Tuebingen, Germany) for zebrafish of the Tuebingen strain.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2005.03.010.

References

[1] A. Valencia, P. Chardin, A. Wittinghofer, C. Sander, The ras protein family: evolutionary tree and role of conserved amino acids, Biochemistry 30 (1991) 4637–4648.
[2] Y. Takai, T. Sasaki, T. Matozaki, Small GTP-binding proteins, Physiol. Rev. 81 (2001) 153–208.
[3] K. Wennerberg, C.J. Der, Rho-family GTPases: it's not only Rac and Rho (and I like it), J. Cell Sci. 117 (2004) 1301–1312.
[4] A.J. Ridley, Rho proteins: linking signaling with membrane trafficking, Traffic 2 (2001) 303–310.
[5] M.C. Kluger, M. Gerhard, A. Schnelzer, K. Borzym, R. Reinhardt, M. Schmitt, E. Lengyel, Isolation and characterization of Rac1 pseudogenes (ψ1Rac1–ψ4Rac1) in the human genome, Gene 341 (2004) 189–197.
[6] F. Rivero, H. Dislich, G. Glockner, A.A. Noegel, The Dictyostelium discoideum family of Rho-related proteins, Nucleic Acids Res. 29 (2001) 1068–1079.
[7] A. Fransson, A. Ruusala, P. Aspenstrom, Atypical Rho GTPases have roles in mitochondrial homeostasis and apoptosis, J. Biol. Chem. 278 (2003) 6495–6502.
[8] R.L. Frederick, J.M. Mc Caffery, K.W. Cunningham, K. Okamoto, J.M. Shaw, Yeast Miro GTPase, Gem1p, regulates mitochondrial morphology via a novel pathway, J. Cell Biol. 167 (2004) 87–98.
[9] W. Tao, D. Pennica, L. Xu, R.F. Kalejta, A.J. Levine, *Wrch-1*, a novel member of the Rho gene family that is regulated by Wnt-1, Genes Dev. 15 (2001) 1796–1807.
[10] K. Nakai, Protein sorting signals and prediction of subcellular localization, Adv. Protein Chem. 54 (2000) 277–344.
[11] S. Ellis, H. Mellor, The novel Rho-family GTPase rif regulates coordinated actin-based membrane rearrangements, Curr. Biol. 10 (2000) 1387–1390.
[12] C.D. Nobes, I. Lauritzen, M.G. Mattei, S. Paris, A. Hall, P. Chardin, A new member of the Rho family, Rnd1, promotes disassembly of actin filament structures and loss of cell adhesion, J. Cell Biol. 141 (1998) 187–197.
[13] X. Li, X. Bu, B. Lu, H. Avraham, R.A. Flavell, B. Lim, The hematopoiesis-specific GTP-binding protein RhoH is GTPase deficient and modulates activities of other Rho GTPases by an inhibitory function, Mol. Cell. Biol. 22 (2002) 1158–1171.
[14] A.H. Meijer, G.S.F. Krens, R.I. Medina, S. He, W. Bitter, E.B. Snaar-Jagalska, H.P. Spaink, Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish, Mol. Immunol. 40 (2004) 773–783.
[15] A.H. Meijer, F.J. Verbeek, E. Salas-Vidal, M. Corredor-Adames, J. Bussman, A. van der Sar, G.W. Otto, R. Geisler, H.P. Spaink, Transcriptome profiling of adult zebrafish at the late stage of chronic tuberculosis due to Mycobacterium marinum infection, Mol. Immunol. In press. doi:10.1016/j.molimm.2004.11.014
[16] D.A. Powers, Evolutionary genetics of fish, Adv. Genet. 29 (1991) 119–228.
[17] D.R. Natale, A.J. Watson, Rac-1 and IQGAP are potential regulators of E-cadherin–catenin interactions during murine preimplantation development, Mech. Dev. 119 (Suppl. 1) (2002) S21–S26.
[18] E. Salas-Vidal, H. Lomeli, Imaging filopodia dynamics in the mouse blastocyst, Dev. Biol. 265 (2004) 75–89.
[19] K. Sugihara, N. Nakatsuji, K. Nakamura, K. Nakao, R. Hashimoto, H. Otani, H. Sakagami, H. Kondo, S. Nozawa, A. Aiba, M. Katsuki, Rac1 is required for the formation of three germ layers during gastrulation, Oncogene 17 (1998) 3427–3433.
[20] R. Habas, I.B. Dawid, X. He, Coactivation of Rac and Rho by Wnt/Frizzled signaling is required for vertebrate gastrulation, Genes Dev. 17 (2003) 295–309.
[21] F. Chen, L. Ma, M.C. Parrini, X. Mao, M. Lopez, C. Wu, P.W. Marks, L. Davidson, D.J. Kwiatkowski, T. Kirchhausen, S.H. Orkin, F.S. Rosen, B.J. Mayer, M.W. Kirschner, F.W. Alt, Cdc42 is required for PIP(2)-induced actin polymerization and early development but not for cell viability, Curr. Biol. 10 (2000) 758–765.
[22] F.J. Sangari, J. Goodman, L.E. Bermudez, Mycobacterium avium enters intestinal epithelial cells through the apical membrane, but not by the basolateral surface, activates small GTPase Rho and, once within epithelial cells, expresses an invasive phenotype, Cell. Microbiol. 2 (2000) 561–568.
[23] C.V. Le, S. Carreno, A. Moisand, C. Bordier, I. Maridonneau-Parini, Complement receptor 3 (CD11b/CD18) mediates type I and type II phagocytosis during nonopsonic and opsonic phagocytosis, respectively, J. Immunol. 169 (2002) 2003–2009.
[24] W.D. Hardt, L.M. Chen, K.E. Schuebel, X.R. Bustelo, J.E. Galan, S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells, Cell 93 (1998) 815–826.
[25] C. Gauthier-Rouviere, E. Vignal, M. Meriane, P. Roux, P. Montcourier, P. Fort, RhoG GTPase controls a pathway that independently activates Rac1 and Cdc42Hs, Mol. Biol. Cell 9 (1998) 1379–1394.
[26] E. Vigorito, S. Bell, B.J. Hebeis, H. Reynolds, S. McAdam, P.C. Emson, A. McKenzie, M. Turner, Immunological function in mice lacking the Rac-related GTPase RhoG, Mol. Cell. Biol. 24 (2004) 719–729.
[27] J. Schultz, F. Milpetz, P. Bork, C.P. Ponting, SMART, a simple modular architecture research tool: identification of signaling domains, Proc. Natl. Acad. Sci. USA 95 (1998) 5857–5864.
[28] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[29] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[30] H.A. Schmidt, K. Strimmer, M. Vingron, A. von Haeseler, TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing, Bioinformatics 18 (2002) 502–504.
[31] B.G. Hall, Phylogenetic Trees Made Easy, Sinauer Associates, Sunderland, MA, 2001.