The *Dictyostelium discoideum* family of Rho-related proteins

Francisco Rivero*, Heidrun Dislich, Gernot Glöckner[1] and Angelika A. Noegel

Institut für Biochemie I, Medizinische Fakultät, Universität zu Köln, Joseph-Stelzmann-Strasse 52, D-50931 Köln, Germany and [1]Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, D-07745 Jena, Germany

ABSTRACT

Taking advantage of the ongoing *Dictyostelium* genome sequencing project, we have assembled >73 kb of genomic DNA in 15 contigs harbouring 15 genes and one pseudogene of Rho-related proteins. Comparison with EST sequences revealed that every gene is interrupted by at least one and up to four introns. For *racC* extensive alternative splicing was identified. Northern blot analysis showed that mRNAs for *racA*, *racE*, *racG*, *racH* and *racI* were present at all stages of development, whereas *racJ* and *racL* were expressed only at late stages. Amino acid sequences have been analysed in the context of Rho-related proteins of other organisms. Rac1a/1b/1c, RacF1/F2 and to a lesser extent RacB and the GTPase domain of RacA can be grouped in the Rac subfamily. None of the additional *Dictyostelium* Rho-related proteins belongs to any of the well-defined subfamilies, like Rac, Cdc42 or Rho. RacD and RacA are unique in that they lack the prenylation motif characteristic of Rho proteins. RacD possesses a 50 residue C-terminal extension and RacA a 400 residue C-terminal extension that contains a proline-rich region, two BTB domains and a novel C-terminal domain. We have also identified homologues for RacA in *Drosophila* and mammals, thus defining a new subfamily of Rho proteins, RhoBTB.

INTRODUCTION

The small GTPases of the Rho family constitute a subgroup of GTP-binding proteins of the Ras superfamily ubiquitously present in eukaryotic cells. The small GTPases act as molecular switches, cycling between an active GTP-bound state and an inactive GDP-bound state, a process that is regulated by GEFs (guanine nucleotide exchange factors) and GAPs (GTPase activating proteins). GEFs catalyse the conversion to the GTP-bound state and GAPs accelerate the intrinsic rate of hydrolysis of bound GTP to GDP. Additionally, GDIs (GDP-dissociation inhibitors) have been described that capture Rho in both GTP- and GDP-bound states and allow it to cycle between cytosol and membranes. In their active state Rho GTPases interact with a multitude of effectors that relay upstream signals to cytoskeletal components, eliciting rearrangements of the actin cytoskeleton (1).

In addition to directly controlling actin reorganisation, but closely related to this activity, Rho GTPases are involved in a diverse range of cellular processes, such as vesicle trafficking, morphogenesis, neutrophil activation, phagocytosis and activation of the NADPH oxidase, mitogenesis, transformation and transcriptional activation (2). In plants, Rho-related proteins are involved in diverse signalling pathways like tip growth, pathogen defense, secondary wall formation and meristem signalling (3). In yeast Rho proteins are involved in cell wall synthesis, control of cell polarity and budding (4,5). A role for Rho GTPases in diverse developmental processes in *Drosophila* and *Caenorhabditis elegans* has been established (6).

Dictyostelium discoideum is an attractive model organism to investigate the components of the actin cytoskeleton and the elements involved in their complex regulatory pathways (7). Indeed, despite their apparent simplicity, *Dictyostelium* amoebae are equipped with a complex actin cytoskeleton that endows the cells with motile behaviour comparable to that of leukocytes. In *Dictyostelium*, 14 Rho-related proteins have been previously identified and have been named Rac1a, Rac1b, Rac1c and RacA–J (8–10), but only a few of them have been characterised. RacE appears to be essential for cytokinesis, but is not involved in processes such as phagocytosis, chemotaxis and development (9,11). A role for RacC in actin cytoskeleton organisation, pinocytosis and phagocytosis has been proposed, based on a study carried out with overexpressor cell lines (12). RacF1 localises to early phagosomes and macropinosomes, but inactivation of the *racF1* gene does not impair endocytosis and other actin-dependent processes, probably due to the presence of a closely related RacF2 (10). Finally, GTPases of the Rac1 group are involved in chemotaxis, cell motility, endocytosis, cytokinesis and development (13,14).

The *Dictyostelium* genome can be easily manipulated by means of recombinant DNA techniques. Since the organism is haploid, mutants can be immediately obtained by homologous recombination, and mutated genes can be introduced with either integrating or non-integrating plasmids. The ongoing *Dictyostelium* genome and cDNA sequencing projects offer a unique opportunity to exploit the advantages of *Dictyostelium* to characterise the signal transduction pathways regulated by

*To whom correspondence should be addressed. Tel: +49 221 478 6987; Fax: +49 221 478 3660; Email: francisco.rivero@uni-koeln.de

Rho GTPases. The *Dictyostelium* genome consists of 34 Mb carried on six chromosomes, plus a multicopy 90 kb extrachromosomal element that harbours the rRNA genes. Using a light shot-gun of the complete genome followed by shot-gun sequencing of individual chromosomes separated by pulsed-field gel electrophoresis (15), the genome project has generated to date >140 Mb sequence information. Complementary to this, sequencing of developmental and sexual cDNA libraries has yielded 3500 non-redundant ESTs (16). We have made extensive use of this information in order to identify the complete set of genes coding for proteins of the Rho family. In this way we could define the entire genomic structure of the previously known 14 members of the family. Furthermore, we found one additional functional gene and one pseudogene. We have deduced some evolutionary implications from sequence comparisons among *Dictyostelium* Rho proteins and with Rho proteins of other species.

MATERIALS AND METHODS

Database searches

We used the following approach to identify genes coding for Rho-related proteins in *D.discoideum*. Initially, BLASTN searches (17) of the *Dictyostelium* EST database (http://www.csm.biol.tsukuba.ac.jp/cDNAproject.html) were performed with the coding sequence of every known *rac* gene as the query, and a comprehensive list of EST clones for each *rac* gene was generated. This allowed us to retrieve sequences of EST clones for all members of the *rac* family except *racF1*, and in most cases to identify cDNA sequences extending beyond the coding region.

In a second step, DNA and amino acid sequences were used for BLASTN and TBLASTN searches, respectively, of the *Dictyostelium* genome project database (http://www.uni-koeln.de/dictyostelium/). For *racE* and *racF1* we used the available genomic DNA sequences (U41222 and AF037042, respectively). Because the genomic database contains raw reads, we set a threshold of >95% identity to consider two sequences as identical. We listed the reads corresponding to every known *rac* gene. This left us with reads for two novel related genes, Ψ*racK* and *racL*. Selected clones were retrieved from the *Dictyostelium* genome project and were fully sequenced using vector-derived and also, when necessary, specific primers. The sequence obtained was used for new screenings of the genomic DNA database in order to complete information on the *rac* genes and their flanking regions.

Searches for Rho proteins of species other than *Dictyostelium* were done using two approaches. First, by means of the Entrez search tool (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi) using keywords like rho, rac and cdc42. Second, using the TBLASTN algorithms in the advanced BLAST option in the non-redundant and EST databases with sequences of several Rho GTPases as query. A similar approach was used to screen the available databases for members of the RhoBTB subfamily with the C-terminal domain of these proteins as the query.

Finally, the deduced amino acid sequence of open reading frames (ORFs) identified upstream and downstream of *rac* genes was used as the query to identify EST clones in the *Dictyostelium* cDNA database, and next to identify homologues in other organisms by means of the TBLASTN algorithm in the non-redundant and EST databases.

Sequence analysis and alignment

Sequence assembly and analysis were performed with the Wisconsin Package Version 9.0 of the Genetics Computer Group (Madison, WI). Protein sequences were aligned using the ClustalX (18) program with a BLOSUM62 matrix and default settings, followed by manual edition with the Bioedit program (19). Sequences of RhoBTB proteins were further analysed for motif or domain composition using the SMART tool (http://smart.embl-heidelberg.de/) and the ProfileScan Server (http://www.isrec.isb-sib.ch/software/PFSCAN_form.html).

Phylogenetic analyses

Accession numbers of the sequences retrieved for the phylogenetic analysis are as follows. *Homo sapiens*: RhoA, L25080; RhoB, X06820; RhoC, L25081; RhoD, O00212; RhoE/Rho8/Rnd3, P52199; Rnd1/Rho6, Q92730; Rnd2/Rho7, P52198; RhoG, NM_001665; Rac1, M29870 (the alternatively spliced variant Rac1b was not included); Rac2, NM_002872; Rac3, AF00859; Cdc42, M57298; GK25, M35543; TC10, M3147; TTF/RhoH, Z35227; Rif, AF239923; RhoBTB1, KIAA0740; RhoBTB2, KIAA0717; RhoBTB3, KIAA0878. *Caenorhabditis elegans*: Cdc42, L10078; Rac1/CED-10, X68492; Rac2, U55018; RhoA, L36965; MIG-2, U82288. *Arabidopsis thaliana*: U88402; Arac11/Rop1, U49971; Arac1/ATGP2, U41295; Arac2, U43026; Arac3/Rop6, U43501; Arac4/Rop2, U45236; Arac5/Rop4, U52350; Arac6/RAC2, AF079487; Arac7, AF079484; Arac8, AF079486; Arac9, AF156896; Arac10, AF079485; ATGP3, U64920. *Entamoeba histolytica*: RacA, U29720; RacB, U29721; RacC, U29722; RacD, U30148; RacG, AF055340; Rho1, L03809. *Gallus gallus*: Cdc42, U40848; Rac1a, U79755; Rac1b, U79756; RhoA, U79757; RhoB, AF098515; RhoC, AF098514. *Saccharomyces cerevisiae*: Cdc42, X51906; Rho1, M15189; Rho2, M15190; Rho3, Q00245; Rho4, Q00246; Rho5, 632414. *Drosophila melanogaster*: Cdc42, U11824; Rac1/RacA, U11823; Rac2/RacB, L38318; Rac3/RhoL, BAA87881; RhoA/Rho1, AF177871; Mt1, AF238044; RhoBTB, AF217287.

For the phylogenetic analyses, only the GTPase core, devoid of hypervariable N- and C-terminal sequences, was considered. The GTPase domain of HsRhoBTB3 is strongly divergent and could not be aligned reliably; therefore it was not included in the analysis. Phylogenetic trees were constructed using the neighbour-joining algorithms (20) of the ClustalX program with correction for multiple substitutions, or the parsimony method of the PHYLIP package (http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html). We did not exclude positions with gaps in order to conserve the information provided by short insertions and deletions present in particular subfamilies of Rho proteins. Bootstrap analysis was applied to provide confidence levels for the tree topology. Construction of trees was done with TreeView (21).

Expression analyses

The expression pattern of *racA*, *racE*, *racG*, *racH*, *racI*, *racJ* and *racL* through the developmental cycle of *Dictyostelium* was analysed by northern blot. *Dictyostelium* amoebae were starved on nitrocellulose filters (22). At indicated time points
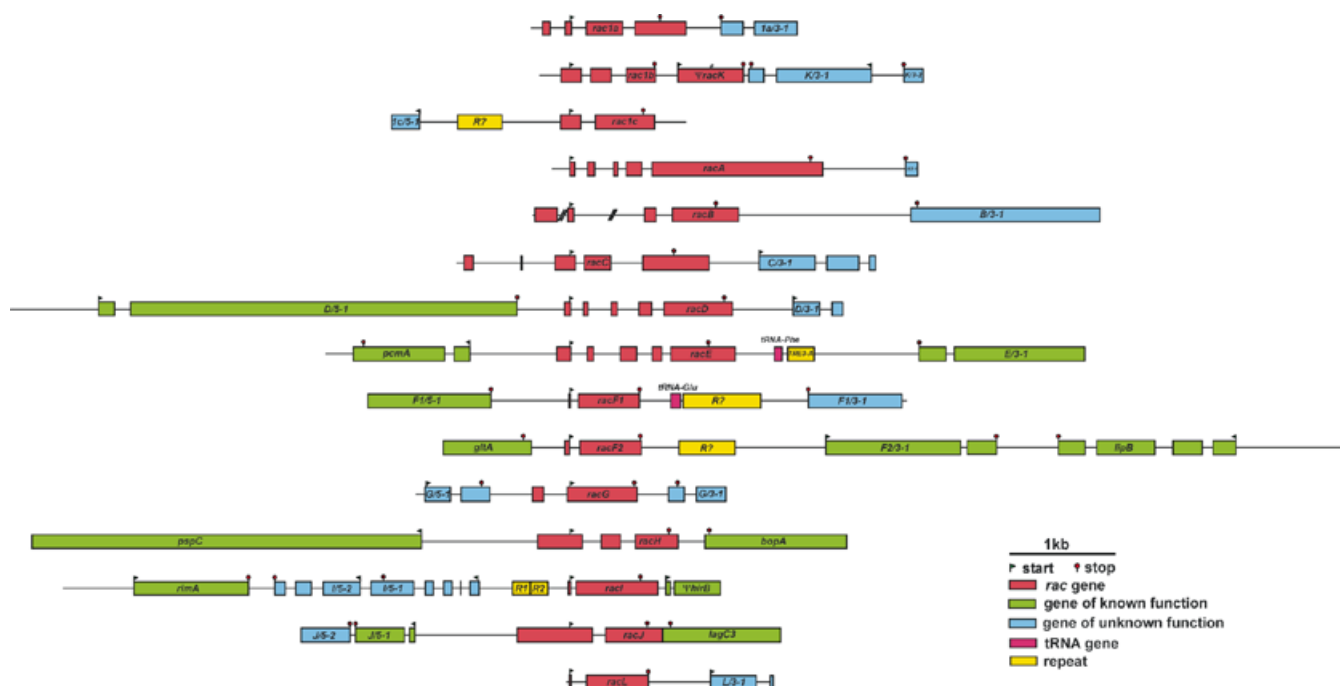
Figure 1. Genomic structure of the *Dictyostelium rac* family. 73.1 kb genomic DNA has been assembled in 15 contigs. Contigs are aligned at the start codon of *rac* genes. Bars indicate coding sequences and are extended beyond the start and stop codon to indicate untranslated but transcribed sequences, when supported by EST clones. Genes other than *rac* genes have been named according to their position relative to the *rac* gene of the corresponding contig (see Table S1 in Supplementary Material for a description). Lines indicate intergenic and intron sequences. When more than one termination site was identified, the most distant has been depicted. A frameshift caused by a 5 nt deletion in Ψ*racK* is indicated by #. Two introns of *racB* are not covered by genomic clones (indicated by slashes). Less than half of the protein ORFs identified correspond to genes whose function is known in *Dictyostelium* or in other species. Accession numbers for the contigs: *rac1a*, AF309947; *rac1b*, AF310884; *rac1c*, AF310885; *racA*, AF310886; *racB*, AF310887; *racC*, AF310888; *racD*, AF310889; *racE*, AF310890; *racF1*, AF037042; *racF2*, AF310892; *racG*, AF310893; *racH*, AF310894; *racI*, AF310895; *racJ*, AF310896; *racL*, AF310897. The contig for *racE* extends sequence U41222. The contig for *racH* overlaps with sequence AF104350, which lacks the *racH* gene. The contig for *racF2* overlaps with sequence AF019108.

samples were taken for RNA extraction as described (23). Samples were transferred onto nylon membranes (Biodyne B, Pall Filtron, Dreieich, Germany) and blots were stained with 0.02% methylene blue in 0.3 M sodium acetate to verify loading. RNA blots containing 20 μg of RNA per lane were incubated with ^{32}P probes (see below) generated using a random prime labelling kit (Stratagene, La Jolla, CA). Hybridisation conditions were as described previously (23).

Molecular biology methods

Standard molecular biology methods were as described by Sambrook *et al.* (24). DNA fragments used for labelling were obtained either by RT–PCR (for *racE*, *racH*, *racI* and *racJ*) or by performing PCR on genomic DNA (for *racA*, *racG* and *racL*). For *racA* a PCR fragment was amplified that corresponded to the C-terminal region of the protein, far from the GTPase domain. For *racL* a forward primer was designed to bypass sequences from the intron located close to the initiation methionine codon. For RT–PCR, first strand cDNA synthesis was performed with MMLV reverse transcriptase (Promega, Madison, WI) on poly(A)$^+$ mRNA purified with the Oligotex system (Qiagen, Hilden, Germany) from total RNA obtained from vegetative cells or after 15 h of development on nitrocellulose. The start codon of *racI* was identified using 5′ RACE with a kit from Roche Diagnostics (Mannheim, Germany) and specific primers. All PCR fragments were cloned into the pGEM-T Easy vector system (Promega) and

sequenced. DNA sequencing was done at the service laboratory of the Center for Molecular Medicine, Cologne, using an automated sequencer (ABI 377 PRISM, Perkin Elmer, Norwalk, CO).

RESULTS

Identification of *Dictyostelium rac* genes and determination of their genomic structure

Following the strategy described in Materials and Methods, we determined the genomic structure of 15 *rac* genes, one of which, *racL*, is a novel gene and one a pseudogene (Ψ*racK*). In most cases this includes intergenic sequences both upstream and downstream of a given *rac* gene (Fig. 1). We decided to keep the nomenclature of previous publications (8–10) and give correlative names to the novel genes Ψ*racK* and *racL*. Ψ*racK* is located immediately downstream of *rac1b* and very close to *K/3-1*, and contains a deletion of 5 nt, as deduced by comparison with other *rac* genes, that causes a frameshift. Pseudogenes for Rho proteins have also been described in *E.histolytica* (25). In total, >73 kb of genomic DNA have been assembled in 15 contigs ranging between 2.1 and 8.4 kb. Of particular interest is the contig containing *racH*. The genes located upstream (*pspC*) and downstream (*bopA*) of *racH* are present in GenBank as a single continuous entry (AF104350) devoid of *racH* sequences. Southern blot analysis on *Dictyostelium* AX2 genomic DNA using a probe encoding

racH was compatible with the expected pattern of bands deduced from the sequences of the *racH* contig. We attribute this discrepancy to strain differences, because the sequence reported in AF104350 was obtained from DH1, whereas the strain chosen for sequencing by the genome project is AX4, a strain nearly identical to AX2.

At the current state of the genome-sequencing project, ascription of a gene to a particular chromosome cannot be made with accuracy. This is due to limitations in the technique used to separate the individual chromosomes. Libraries for a particular chromosome are contaminated to 50% with sequences from all other chromosomes. However, based on the frequency of reads from a particular library for a given contig, its localisation can be estimated in some cases, for example for *rac1b*, *racF1* and *racH* (chromosome 1), *racB*, *racE* and *racF2* (chromosome 2) and *racJ* (chromosome 6). Except for *rac1b* and Ψ*racK*, *Dictyostelium rac* genes do not appear to form clusters in the genome, at least in the range of our contigs.

Comparison of genomic and EST sequences was used to identify and analyse introns and termination regions. Due to the nature of the library sequenced by the *Dictyostelium* cDNA project, with clones of variable length at the 5′ extreme that in some cases do not reach the start codon, no attempt has been made to identify putative transcription start sites. For the same reason, additional introns located upstream of the coding region of some genes could have been missed. For *racF1* and *racL*, for which EST clones are not available, introns were deduced by comparison with other *rac* genes. Every *rac* gene is interrupted by at least one and up to four introns. In *rac1a*, *racB*, *racC* and *racG* there is at least one intron located upstream of the start codon. The only *rac* gene not interrupted by introns in the coding region is *racG*. The presence of so many introns contrasts remarkably with the relative paucity of introns of most *Dictyostelium* genes, as is evident from a comparison with adjacent genes in Figure 1.

We have identified and annotated protein ORFs, tRNA genes and repetitive elements present upstream and downstream of *rac* genes. This information is depicted in Figure 1 and described in Table S1 (see Supplementary Material). The average length of the intergenic regions in our sample, as calculated from start to start, start to stop or stop to stop codons, was 727 nt. This distance is much shorter if only untranscribed regions are considered. For example, the transcripts of *racJ* and *lagC3* overlap at their 3′ ends (Fig. 1), leaving virtually no intergenic region. Intergenic regions are larger (1.2 kb on average) between start codons, indicating a higher requirement of sequence length for the transcription machinery and the regulatory elements at the promoter regions.

Expression of *rac* genes throughout development

One prominent feature of the *Dictyostelium* life cycle is the transition from single-cell amoebae to a multicellular fruiting body consisting of at least two differentiated cell types. This transition is triggered by starvation of the cells and involves coordinated transcription of certain genes and differentiation and sorting out of cell populations. A previous report had shown complex patterns of developmental regulation for *rac1a/1b/1c*, *racB*, *racC* and *racD* (8). We have used northern blot analysis to study the expression of the remaining *rac* genes during synchronised development on nitrocellulose filters. In addition, knowledge of the full-length sequence of *racA*
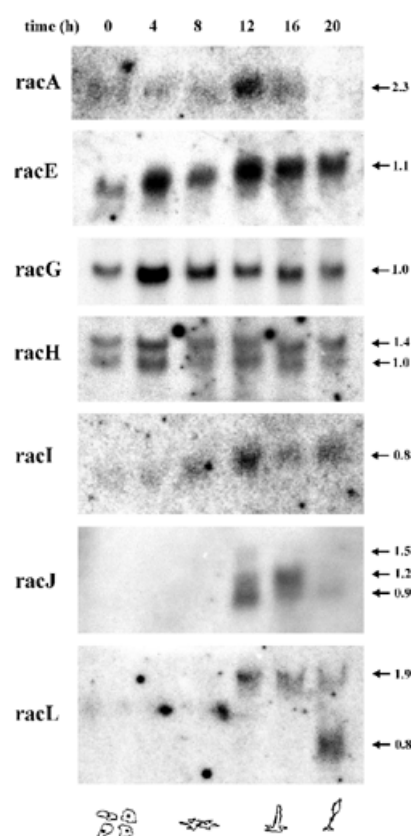

Figure 2. Patterns of mRNA expression of *rac* genes during the developmental cycle. *Dictyostelium* cells were allowed to develop on nitrocellulose filters and total RNA was extracted at the indicated time points. Northern blots containing 20 µg total RNA for each time point were hybridised with the corresponding ^{32}P-radiolabelled probe. Two different blots were used; for *racA*, *racE*, *racI*, *racJ* and *racL* the same blot was used; for *racG* and *racH* a separate blot was used. The lane corresponding to 4 h was slightly overloaded. The blots were stripped and tested for absence of radioactive signals prior to hybridisation with a new probe. The arrows mark the size of the transcripts in kilobases. The developmental stages corresponding to the time points are depicted below. The transcripts of some of the *rac* genes analysed here, most notably *racJ* and *racL*, are present after the aggregation stage. Very long exposure times were required in order to detect *racA*, *racI*, *racJ* and *racL*.

prompted us to repeat analysis of the expression pattern of this gene. For this gene Bush *et al.* (8) reported a transcript of 0.9 kb, clearly not compatible with the 1.8 kb of coding sequences described here. This result was probably due to cross-hybridisation with another *rac* gene; therefore, we used sequences derived from the 3′ region of *racA* in the northern blot analysis.

Rac genes display diverse patterns of developmental regulation. Genes like *racA* and *racI* are very weakly expressed at all stages and display maximum levels after 12 h of starvation, corresponding to the first finger stage (Fig. 2). In contrast, *racE*, *racG* and *racH* are present at high levels throughout the developmental cycle. *RacE* levels increase at late developmental stages, with a maximum at 12 h. Finally, *racJ* and *racL* are of special interest, being expressed at low levels and exclusively after 12 h of starvation, when culmination and maturation of the fruiting body take place. For *racH*, *racJ* and *racL* we have observed two or more transcripts of different
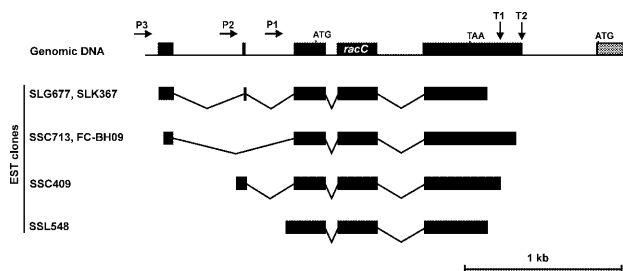
Figure 3. Alternative splicing of *racC*. Comparison of genomic and EST sequences allowed the identification of extensive alternative splicing in the *racC* gene. At least three promoters and two terminators account for the presence of diverse EST species (only informative EST clones are presented).

sizes. This has been reported previously for *rac1a/1b/1c* and *racD* (8) and for *racF1/F2* (10).

Comparison of genomic sequences with sequences of the available EST clones indicates the presence of at least three promoter regions and two termination regions in *racC* (Fig. 3). Alternative splicing takes place around one 764 nt intron placed upstream of the start codon. When transcription starts in the promoter region P3, this intron is completely spliced out as in EST clones SSC713 and FC-BH09, and can be spliced into two smaller sub-introns, leaving an intervening exon of only 12 nt, as in EST clones SLG677 and SLK367. Interestingly, when the entire intron is spliced, the donor site is placed 6 nt upstream of the donor site used for splicing in two parts. The promoter region P2 is apparently located in the first sub-intron and transcription from this region (EST clone SSC409) leads to splicing of the second sub-intron at exactly the same sites as in clones SLG677 and SLK367. Finally, the promoter region P1 is apparently located in the second sub-intron (clone SSL548). With the available data we cannot discriminate whether a particular combination of promoter and terminator is preferred. The transcripts arising from all possible combinations would range between ~1000 and 1150 nt, and are therefore difficult to resolve in northern analysis. In fact, published northern blot analyses show a broad band of compatible size (8). In any case, the coding region of *racC* remains unaltered by the transcriptional events described above.

Conserved sequence elements of *Dictyostelium* Rho GTPases

To better appreciate the relationships among the members of the *Dictyostelium* Rho family and to analyse the requirements for their function as Rho GTPase, we have generated a multiple alignment (Fig. 4). These relationships can also be appreciated in the phylogenetic tree (Fig. 6). With the exception of RacA (described below in more detail), RacD and RacE, all *Dictyostelium* Rac proteins are 192–205 residues long. RacD and RacE possess serine-rich insertions of different lengths immediately preceding the membrane association domain. RacF1, RacF2, RacA and RacB are the most closely related to Rac1a/1b/1c. Neither of the other Rac proteins is appreciably related to each other, RacI and RacJ being the most divergent. Despite the internal deletion of ΨRacK, the reconstructed amino acid sequence of this pseudogene is fairly well conserved, particularly in some important functional regions involved in GTPase activity and membrane association, indicating that Ψ*racK* arose very recently.

We have analysed five conserved regions of proteins of the Ras superfamily involved in nucleotide and magnesium ion binding (26,27). All *Dictyostelium* Rac proteins conform to the consensus of the phosphate binding loop L1 (P-loop or P/M1). Only RacA and RacJ are special in that the magnesium ion-binding threonine, characteristic of Rho proteins, is substituted by a serine, more characteristic of Ras and some Rab proteins. In RacJ the region between α-helix A1 and β-strand B4, which encompasses both switch I and II, is strongly divergent from the other Rac proteins, and the conserved elements present in this region cannot be reliably identified. In the switch I region, involved in GTP hydrolysis, binding to GAPs and downstream signal transduction, the conserved phenylalanine of G1 is replaced by isoleucine in RacG, and the conserved threonine of P/M2 is replaced by serine in RacI. Of the element P/M3, the invariant aspartate involved in magnesium ion binding is conserved in all Rac proteins, and the invariant glycine involved in γ-phosphate binding is replaced by alanine in RacI. The guanine specificity region (G2) is conserved in all *Dictyostelium* Rac proteins, with the exception of RacG and RacL, which have the sequence TQXD; all other proteins present the signature TKXD in this region, characteristic of the Rac subfamily. Finally, in the G3 region, which helps in the binding of the guanine base, RacI (SSL), RacJ (STA) and RacL (SVV) present some deviations from the consensus SA(K/L).

The most characteristic signature of Rho proteins is the Rho insert, a 13 amino acid insertion between β-strand B5 and α-helix A4. In HsRac1 (28) and HsRhoA (29) this insertion consists of one α-helix (A3′) followed by an extended loop that constitutes a mobile and exposed region with a highly charged surface. The Rho insert is one of the regions determining the specificity of functions of Rac against GTPases of other families. All *Dictyostelium* Rac proteins present a Rho insert rich in charged residues, although in some (RacA, RacE, RacH and RacJ) this insert is shorter than 13 amino acids. There is clearly more variability in the sequences of the inserts than in the rest of the protein and indeed, the insert is very variable not only among subfamilies but within members of a subfamily of Rho GTPases as well.

Finally, all Rac proteins except RacD and RacA end with a CAAX prenylation motif, a signal for attachment of a lipid moiety, geranylgeranyl or farnesyl, characteristic of Rho proteins. A polybasic domain rich in lysine residues precedes this motif. Prenylation and the polybasic domain have been demonstrated to contribute to the association of Rho proteins with membranes (30). In RacA the polybasic region is interrupted by a serine-rich stretch and is followed by additional structural domains (see below). In RacD a prominent polybasic region is present, but the prenylation motif is absent.

RacA is a RhoBTB protein

In the course of sequencing of genomic clones for *racA* we found that the incomplete amino acid sequence of RacA that has already been published (8) was not followed by a prenylation motif and a stop codon. Instead, a continuous ORF that codes for approximately 400 additional residues followed. Therefore, the *Dictyostelium racA* gene encodes an unusual protein of 598 residues with a GTPase domain at its N-terminus and a novel C-terminal region. A search for domains or motifs within the 400 residue C-terminal moiety of RacA identified two BTB domains. The BTB domain (Broad-Complex, Tramtrack and
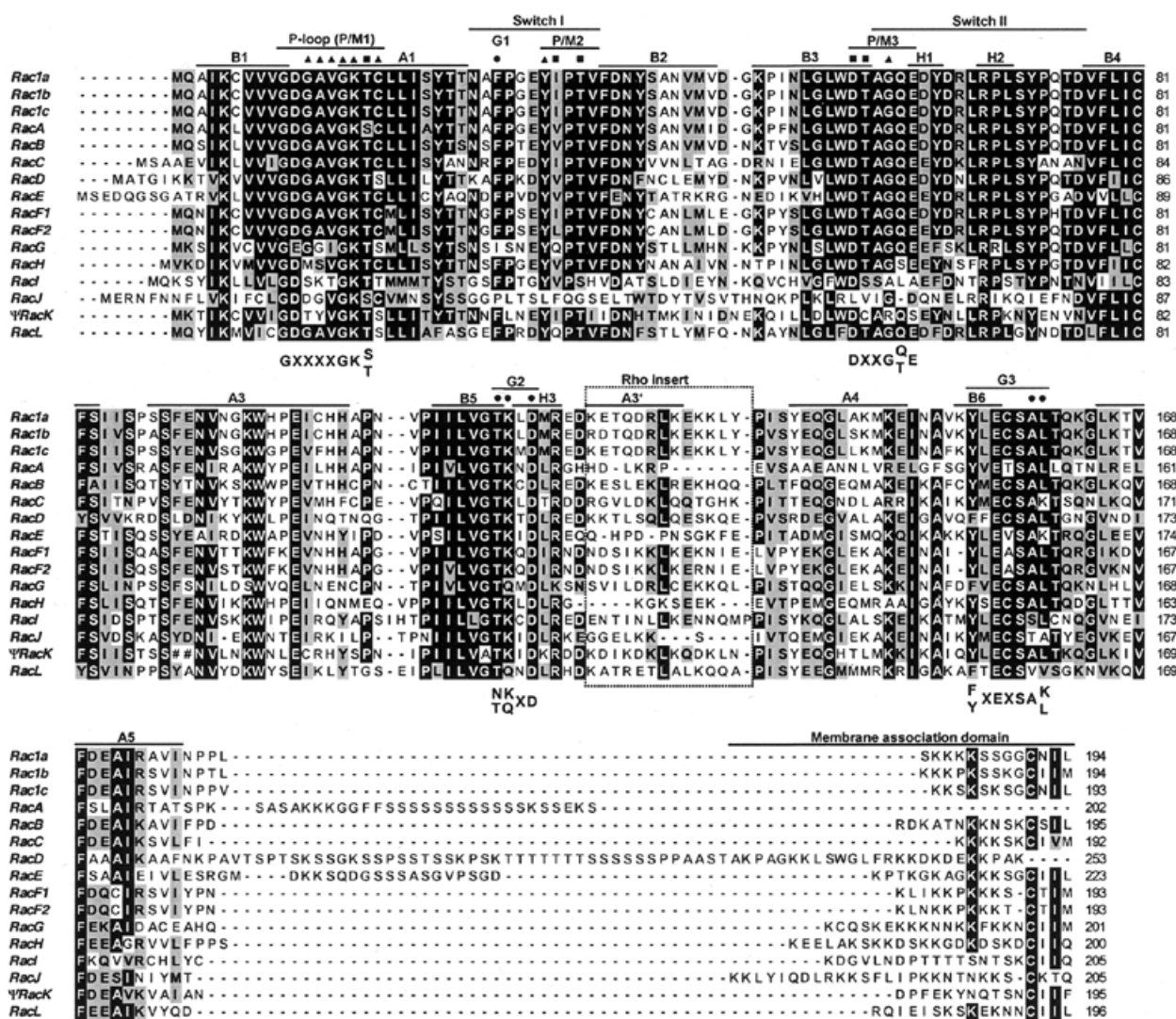
Figure 4. Multiple alignment of *Dictyostelium* Rho-related proteins. Sequences were aligned with ClustalX and default parameters, and the output file was subsequently edited manually. RacA has been truncated at residue 202 to include only the GTPase domain. The deduced amino acid sequence of ΨRacK is also included for comparison (## denotes the deletion of two residues). Dashes indicate gaps introduced for optimal alignment. Residues identical or similar in at least 60% of the sequences are boxed in black or grey, respectively. Secondary structure elements as determined for human RhoA are indicated at the top of the aligned sequences: A1–5, α-helices; B1–6, β-strands; H1–3, 3₁₀-helices. Functional regions are indicated above the structural elements and consensus sequences for some of them are indicated below the aligned sequences. Symbols denote residues involved in phosphate binding (triangles), magnesium ion binding (squares) and nucleotide binding or stabilisation (circles). The Rho insert is boxed.

Bric à brac), also known as POZ domain (poxvirus and zinc finger), is an evolutionarily-conserved domain involved in protein–protein interaction, participating in homomeric and heteromeric associations with other BTB domains (31). The crystal structure of some BTB domains has been solved. They constitute tightly intertwined dimers with an extensive hydrophobic interface. The folding consists of a cluster of α-helices flanked by short β-sheets at both the top and bottom of the molecule (32).

The unusual structure of *Dictyostelium* RacA prompted us to explore the sequence databases with the C-terminal moiety of this protein in search of homologues in other species. A search of the non-redundant database yielded two entries for *D.melanogaster* (AF217287 and AF221547) and three for human (KIAA0740, KIAA0717 and KIAA0878) where similarity

was not restricted to the BTB domains. Inspection of both *Drosophila* sequences and comparison with genomic sequences of this species indicated that one (AF221547) contained several deletions of single nucleotides and was therefore discarded. Sequence AF217287 has been annotated as RhoBTB, therefore we have applied the same nomenclature to the human orthologues and have renamed them RhoBTB1 (KIAA0740), RhoBTB2 (KIAA0717) and RhoBTB3 (KIAA0878). Inspection of genomic sequences present in the database identified the missing first 100 residues of the entry corresponding to RhoBTB2. Finally, a search of the EST database indicated the presence of RhoBTB counterparts in other mammalian species, whereas search of specific databases indicated that neither *S.cerevisiae*, *C.elegans* nor *A.thaliana* appear to harbour RhoBTB orthologues.
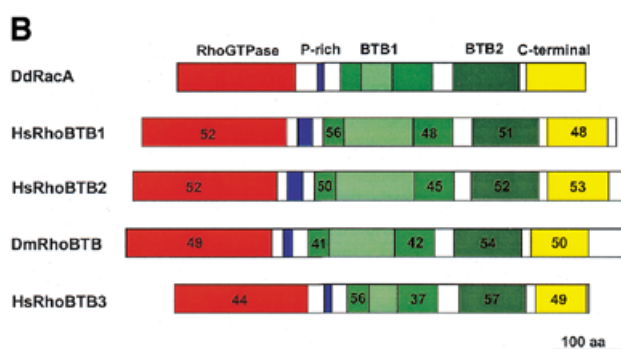
A

(Figure 5A: multiple sequence alignment of DdRacA, HsRhoBTB1, HsRhoBTB2, DmRhoBTB and HsRhoBTB3, with labelled domains: RhoGTPase, Proline-rich, BTB1, BTB2 and C-terminal domain.)

From the alignment of *Dictyostelium* RacA with *Drosophila* and human orthologues it becomes evident that (i) homology is not limited to the novel C-terminal region, but also extends to the GTPase domain, and (ii) additional domains can be recognised on each side of the BTB domains (Fig. 5A). With the exception of HsRhoBTB3, whose GTPase domain is very divergent and cannot be reliably aligned with other Rho GTPases, and in contrast to DdRacA and Rho GTPases in general, the GTPase domain of *Drosophila* and mammalian RhoBTB proteins contains two insertions (six residues between α-helix A1

Figure 5. (Opposite and above) Multiple alignment and sequence comparison of RhoBTB proteins. (**A**) Multiple alignment and domain structure of RhoBTB proteins. A search of GenBank databases with the 400 residues long C-terminal part of DdRacA led to the identification of the complete sequence of one orthologue in *Drosophila* and three in human. Proteins of this novel subfamily are composed of a GTPase domain, a proline-rich region, two BTB domains (the first one divided into two subdomains by intervening sequences; dashed line) and a C-terminal domain of unknown function. Note that only HsRhoBTB3 ends with a prenylation signal. Sequences were aligned with ClustalX. Dashes indicate gaps introduced for optimal alignment. Residues identical or similar in at least three sequences appear in a black or grey background, respectively. Accession numbers are as follows: DmRhoBTB, AF217287; HsRhoBTB1, KIAA0740; HsRhoBTB3, KIAA0717; HsRhoBTB3, KIAA0878. (**B**) Comparison of DdRacA with other RhoBTB proteins. The numbers in each corresponding-colour region represent the percent similarity of each major domain of RacA to any of the other RhoBTB proteins. For the first BTB domain both subdomains have been calculated separately.

and switch I, and 10 residues between β-strands B2 and B3) and one deletion (two residues in switch II immediately after the P/M3 element). Other deviations from the GTPase consensus are the substitution of phenylalanine of element G1 by leucine, the presence of cysteine instead of asparagine or threonine in G2 (similar to members of the Rnd family and other Rho proteins like ScRho1 and HsRhoD) and the presence of SV(V/F) instead of SA(L/K) in G3. A Rho insert, signature of Rho proteins, is also present.

All five RhoBTB proteins share a proline-rich region (most prominent in HsRhoBTB1, HsRhoBTB2 and DmRhoBTB) linking the GTPase to the first BTB domain. This region could act as an SH3 domain binding site. SH3 domains are often present in proteins involved in signal transduction related to cytoskeletal organisation (33). Immediately after the second BTB domain another region of high similarity among all five sequences was identified. This region does not match any sequences present in the domain databases and probably constitutes a novel domain. The last third of this region displays a high content of basic residues and, interestingly, HsRhoBTB3 ends with a prenylation motif. In addition, HsRhoBTB1, HsRhoBTB2 and DmRhoBTB have extensions of variable length at their C-termini. Common to all RhoBTB proteins is also the interruption of the first BTB domain by intervening sequences of variable length that are rich in charged amino acids. These sequences are unrelated among the different RhoBTB proteins. In HsRhoBTB2 this insertion contains a histidine-rich stretch.

According to the modular architecture described above, we have grouped these proteins in a novel subfamily of Rho GTPases that we have named RhoBTB, following the nomenclature already proposed for the *Drosophila* member. Sequence comparisons among the members of this subfamily indicate that HsRhoBTB1 and 2 are closely related to each other (79% similarity) and to DmRhoBTB (56% similarity). DdRacA and HsRhoBTB3 are more divergent, with 41–43% and 37–42% similarity, respectively, to other RhoBTB proteins. The degree of similarity increases when sequence comparisons are restricted to the domains, as is shown for DdRacA in Figure 5B.

Phylogenetic analysis

Mammalian Rho GTPases have been classically grouped into three major subfamilies, Rho, Rac and Cdc42. In fibroblasts, signalling occurs through a hierarchical cascade in which activated Cdc42 activates Rac, which subsequently activates Rho. Cdc42 induces formation of filopodia, Rac induces lamellipodia formation and membrane ruffling and Rho causes formation of stress fibres and focal adhesion complexes (34). In order to obtain clues about potential activities of members of the *Dictyostelium* Rho family, we analysed the amino acid sequence of these proteins in the context of Rho proteins of other organisms. To this end, and to investigate whether different species share groups or subfamilies of Rho GTPases, we constructed a phylogenetic tree based on the alignment of complete sets of sequences of Rho proteins from selected organisms, including representatives of protists, fungi, plants, invertebrates and vertebrates. For *S.cerevisiae*, *C.elegans* and *D.melanogaster*, completion of sequencing of their respective genomes ensures that the complete family of Rho GTPases is represented.

Our analysis shows that, with the exception of DdRac1a/1b/1c, RacF1/F2 and to a lesser extent RacB and RacA, which are close to the Rac subfamily, neither of the additional *Dictyostelium* Rho-related proteins falls into any of the other well-defined subfamilies, particularly Cdc42 or Rho (Fig. 6). Interestingly, the GTPase domain of DdRacA appears more closely related to Rac proteins than to RhoBTB proteins, where RacA actually belongs according to the domain structure of this novel subfamily (Fig. 5). The analysis identified EhRacC and AtU88402 as the closest relatives of DdRacC and DdRacE, respectively.

For many Rho GTPases bootstrap analysis supports their clustering into distinct monophyletic subfamilies like Rac, Cdc42, Rho (in a narrow sense), Rop, Rnd and RhoBTB. However, there are still many proteins, among them eight from
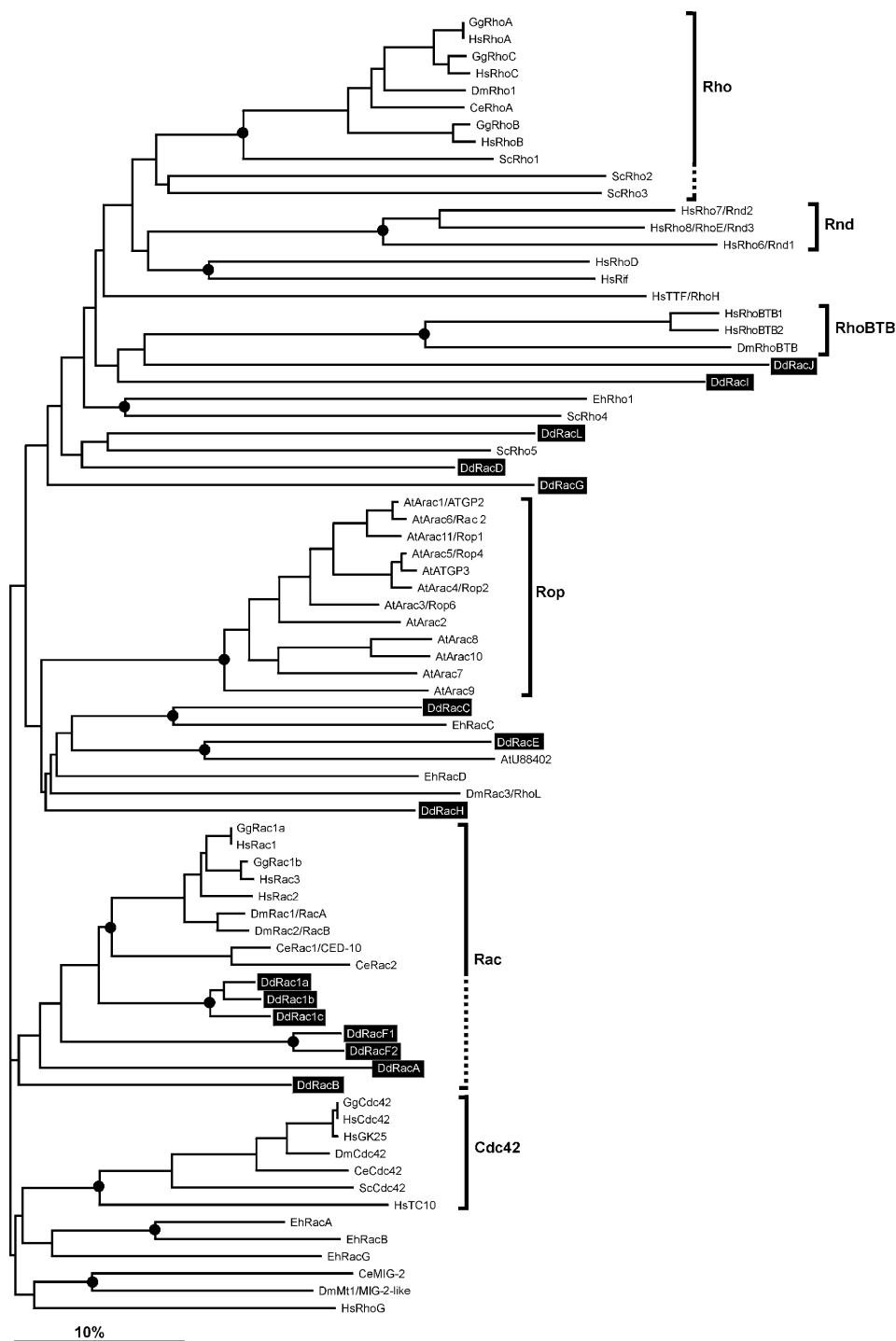
Figure 6. Phylogenetic tree of the Rho family. Representative species for which fairly complete sets of Rho proteins are available have been selected. Besides the 15 *D.discoideum* (Dd) Rho proteins, we also retrieved for this study sequences of six Rho proteins from *E.histolytica* (Eh), six from *S.cerevisiae* (Sc), 13 from *A.thaliana* (At), five from *C.elegans* (Ce), seven from *D.melanogaster* (Dm), six from *G.gallus* (Gg) and 19 from *H.sapiens* (Hs). Amino acid sequences of the GTPase core were aligned and a bootstrapped unrooted phylogenetic tree was constructed using the neighbour joining method as described in Materials and Methods. A consensus tree obtained with the parsimony method yielded comparable results. The innermost node of each branch supported by >50% bootstraps has been marked with a circle. The scale bar indicates 10% divergence. *Dictyostelium* Rho-related proteins are highlighted in black. Well-defined subfamilies have been marked on the right, using dashed lines to include closely related members not clearly supported by bootstrapping. Except for DdRac1a/1b/1c, RacF1/F2 and, to a lesser extent, RacB and RacA, which can be grouped in the Rac subfamily, neither of the additional *Dictyostelium* Rho-related proteins fall into any of the other subfamilies. Note that the DdRacA GTPase domain is more closely related to Rac proteins than to RhoBTB proteins, where RacA actually belongs. The closest relatives of DdRacC and DdRacE are EhRacC and AtU88402, respectively.
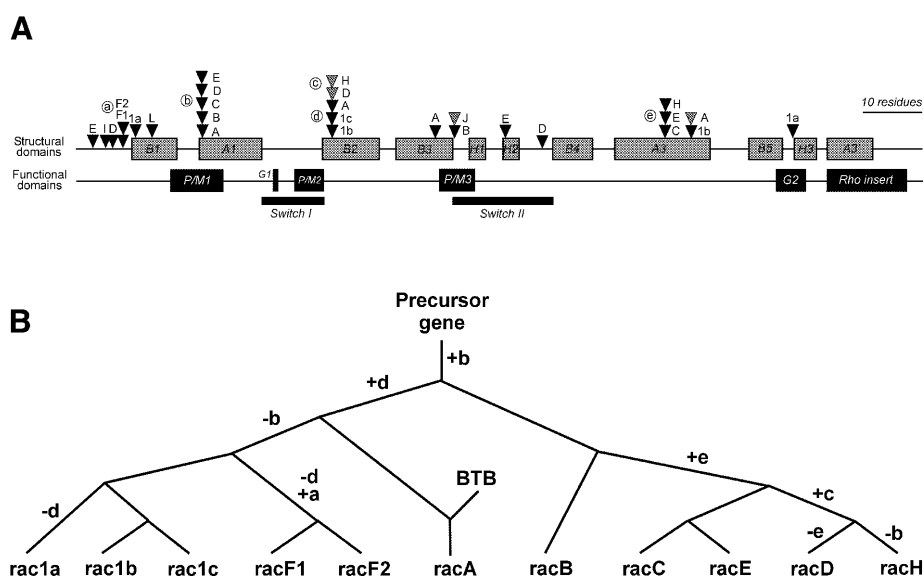
Figure 7. Analysis of intron positions in *Dictyostelium rac* genes. (**A**) Locations of introns relative to regions encoding structural and functional elements in the *rac* genes. Only introns placed in coding regions are shown. There are no introns beyond helix H3. Positions of introns are depicted as triangles. Introns residing in the same codon but at different locations are discriminated by their shading. Positions shared by two or more genes have been labelled with circled letters. (**B**) Evolutionary relationships among some *Dictyostelium rac* genes. The tree is based on the analysis of shared intron positions combined with analyses of protein sequences. Nodes represent duplication events or, for *racA*, fusion with the BTB domain-containing C-terminal moiety. Letters at the branches denote acquired (+) or lost (−) introns. The nomenclature for the intron positions is as in (A). Branch lengths are not proportional to evolutionary distance. The *Dictyostelium rac* family probably arose by a complex process of duplications and acquisition and loss of introns.

Dictyostelium, that radiate from the tree without apparent phylogenetic relatives. Note that DdRacI and DdRacJ are among the most distant members of the Rho family.

Analysis of intron positions of *Dictyostelium rac* genes

To gain insight into the evolutionary history of the *Dictyostelium rac* genes and to investigate a possible correlation of intron location with structural and functional domains of the small GTPases, we have analysed the position and sequence of the introns present in the coding region of all *rac* genes (Fig. 7A). Introns appeared distributed in 18 different positions, preferentially in the 5′-half of the genes, and they could be found to interrupt a codon at any position. In three cases introns were identified at different positions within the same codon. With very few exceptions, we did not observe a preferential distribution of intron positions with respect to regions defining structural or functional domains.

In five cases one position was shared by two or more genes (labelled a–e in Fig. 7A), suggesting that the intron in that position was already present in an ancestral precursor gene. These shared locations may, therefore, be very informative in terms of evolutionary implications. For this reason we attempted to determine the extent of sequence conservation among introns at those locations. Only in a few cases were we able to identify recognisable sequence elements that were shared among introns at the same location, namely positions a, d and e. In addition to high sequence variability, introns of the same position may differ greatly in size; for example introns of *racE* and *racC* at position e are 84 and 315 nt long, respectively. Interestingly, introns of *racA* at positions b and d are more related to each other, in terms of sequence similarity, than to

other introns of the same respective position, indicating that they arose by a duplication event.

DISCUSSION

We have made extensive use of the information released by the ongoing *Dictyostelium* sequencing projects to identify additional members and investigate the genomic organisation of the family of Rho GTPases. In *Dictyostelium* this family is composed of 15 genes and one pseudogene. All genes are transcribed and most of them are represented in the developmental and sexual cDNA libraries sequenced at the University of Tsukuba (16). Currently, >140 Mb of genomic DNA sequence is available. On average, this corresponds to a 4-fold coverage over the whole genome, and based on the amount already sequenced, the probability of having identified all members of the Rho family can be calculated as being >95% (35).

Expression patterns of *Dictyostelium rac* genes

Northern blot analyses indicate that expression of several *rac* genes displays some degree of developmental regulation. According to the expression patterns, three main groups of *rac* genes could be established. Most genes are expressed during the complete developmental cycle, either at high levels, like the *rac1* group (8), *racE*, the *racF* group (10), *racG* and *racH*, or very weakly, like *racA* and *racI*. Genes like *racB*, *racC* and *racD* are characteristic of vegetative and early developmental stages (8), whereas *racJ* and *racL* are expressed only at later stages. A correlation of gene expression pattern with protein levels is missing, but these results are clearly suggestive of

distinct roles for particular Rac proteins at different stages of development.

The presence of more than one mRNA species for almost all *rac* genes in the northern blot analyses (8,10; Fig. 2) can be attributed to the use of alternative promoters, alternative terminators or alternative splicing of introns. An exception appears to be *racG,* for which a single sharp band is detected in the northern blot analysis. This is consistent with the clustering of all informative EST clones downstream of a single poly-adenylation signal. For four *rac* genes, *rac1a*, *racC*, *racE* and *racJ*, our analysis of the available information allowed the identification of more than one cleavage site. Other genes, like *racD* (8), *racF1/F2* (10), *racH* or *racL*, are clearly represented in northern blot analyses by more than one mRNA species, however the sequence information available is insufficient to identify the cause. Only for *racC* have we obtained clear indications of the use of alternative promoters combined with alternative splicing of introns (Fig. 3). The relevance of the complex pattern of gene expression of many of the *rac* genes in terms of spatial and temporal regulation and the factors involved in this regulation remain to be established.

Dictyostelium Rho proteins in the context of other species

Some conclusions can be drawn from our phylogenetic analysis of the Rho family. First, many Rho proteins cannot be grouped into well-defined subfamilies. Although more subfamilies are likely to emerge with the discovery and characterisation of novel members in other organisms, it is also likely that each organism possesses specialised divergent Rho members. Second, not all well-defined subfamilies are shared by all species. Plants, for example, constitute an extreme case because they almost exclusively possess Rop proteins. Both animals and yeast have members of the Rho and the Cdc42 subfamilies, but yeast does not have Rac proteins. In fact, an exhaustive search for Rho (in a narrow sense), Rac and Cdc42 orthologues in the current sequence databases extended this observation to many other animal and fungal species than the ones analysed here. *Dictyostelium*, like animals, has representatives of the Rac subfamily, but lacks Rho and Cdc42 proteins, and members of any of these three subfamilies have not been identified so far in *Entamoeba*.

This phylogenetic analysis places *Dictyostelium* closer to animals and fungi than to plants, in agreement with data based on the analysis of diverse protein sets (36). Additionally, it supports the view that a considerable amount of diversification of Rho proteins into the contemporary subfamilies occurred very early during evolution, and that this process was accompanied by loss and emergence of new subfamilies. This is particularly evident in the RhoBTB subfamily, which emerged very early, because it is present in *Dictyostelium*, but was apparently lost in *C.elegans*. The existence of multiple paralogues in every species and the high degree of functional redundancy among the Rho GTPases could have arisen during the evolutionary process in which novel subfamilies of Rho proteins took over the control of functions acquired recently by the cells, whereas the functions of extinct members were taken over by other Rho proteins.

Together with the sequence comparisons, the analysis of intron positions might help to elucidate the evolutionary relationships among *Dictyostelium rac* genes. The preferential location in the 5′ half of the *rac* genes and the lack of correlation of the position with borders of structural and functional domains conform to the results of an analysis of introns in 28 small GTPases of the ARF, Rab and Ras families of several species. Neither of the positions of *Dictyostelium* genes matches the positions of human *rac2* or *rhoG* or *C.elegans rac2* reported by Courjal *et al.* (37), and the three positions that match the results of Dietmaier and Fabry (38) are either irrelevant or can be considered coincidental given the wide distribution of introns along the GTPase genes. Taken together, all these analyses confirm that genes of the same family have more common intron positions than when compared with genes of other families, and that within a particular family there exists a broad variability in the intron positions among species. This suggests, in agreement with our interpretation of the data from the phylogenetic analysis, that the radiation of small GTPases of the Ras superfamily into the different families took place very early during evolution, and that the intron–exon structure arose by later multiple events of insertion and loss of introns. This supports the 'introns-late' theory, which postulates that spliceosomal introns were inserted in genes after eukaryotic gene diversification had occurred (39).

Evolution of *Dictyostelium rac* genes

The evolutionary history of a subset of *Dictyostelium rac* genes can be determined from the presence of some intron locations shared by two or more *rac* genes combined with analyses of protein homologies (Fig. 7B). A common ancestor gene was characterised by an intron at position b. This position is probably very ancient because five genes share it. This ancestor underwent duplication and gave rise, on the one hand, to *racB* and, after diverse events of intron insertions and losses, to *racC*, *racE*, *racH* and *racD*. On the other hand, the ancestor gene gained an intron at position d, presumably by duplication of the intron at position b. It cannot be ruled out that the intron at position b arose by duplication of the intron at position d, but even in that case the topology of the resulting evolutionary tree does not differ much from that presented in Figure 7B. The ancestor carrying introns at positions b and d gave rise to *racA* and, after loss of the intron at position b, to the group of *rac1* genes. The predecessor of the *rac1* genes underwent two duplications, yielding *rac1a*, *rac1b* and *rac1c* after additional losses and insertions of introns. Sequence analysis places RacF1/F2 close to the Rac1 group (Fig. 4). It is therefore very likely that the *racF1* and *racF2* genes, which share the intron at position a, are derived from the same ancestor as the *rac1* group after loss of the intron at position d and gain of the intron at position a. Later on, the intron sequences diverged at a higher rate than the surrounding coding sequences. *RacD* underwent additional changes, such as the insertion of a serine and threonine-rich stretch of amino acids and the loss of the prenylation motif. *RacA* probably arose by fusion of a *rac* gene with another gene carrying the proline-rich, BTB and C-terminal domains. This gene poses an interesting problem: both the sequence and the intron analyses place *racA* close to members of the Rac family, yet the GTPase domain of the *Drosophila* and human orthologues is significantly divergent from Rac proteins. It appears that the RhoBTB family diverged from a *rac* gene very early during evolution and underwent additional modifications in higher eukaryotes that did not take place in *Dictyostelium*.

REFERENCES

1. Hall,A. (1998) Rho GTPases and the actin cytoskeleton. *Science*, **279**, 509–514.
2. Van Aelst,L. and D'Souza-Schorey,C. (1997) Rho GTPases and signaling networks. *Genes Dev.*, **11**, 2295–2322.
3. Valster,A.H., Hepler,P.K. and Chernoff,J. (2000) Plant GTPases: the Rhos in bloom. *Trends Cell Biol.*, **10**, 141–146.
4. Arellano,M., Coll,P.M. and Pérez,P. (1999) Rho GTPases in the control of cell morphology, cell polarity and actin localization in fission yeast. *Microsc. Res. Tech.*, **47**, 51–60.
5. Pruyne,D. and Bretscher,A. (2000) Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states. *J. Cell Sci.*, **113**, 365–375.
6. Settleman,J. (1999) Rho GTPases in development. *Prog. Mol. Subcell. Biol.*, **22**, 201–229.
7. Noegel,A.A. and Schleicher,M. (2000) The actin cytoskeleton of *Dictyostelium*: a story told by mutants. *J. Cell Sci.*, **113**, 759–766.
8. Bush,J., Franek,K. and Cardelli,J. (1993) Cloning and characterization of seven novel *Dictyostelium discoideum rac*-related genes belonging to the *rho* family of GTPases. *Gene*, **136**, 61–68.
9. Larochelle,D.A., Vithalani,K.K. and De Lozanne,A. (1996) A novel member of the rho family of small GTP-binding proteins is specifically required for cytokinesis. *Mol. Biol. Cell*, **133**, 1321–1329.
10. Rivero,F., Albrecht,R., Dislich,H., Bracco,E., Graciotti,L., Bozzaro,S. and Noegel,A.A. (1998) RacF1, a novel member of the Rho protein family in *Dictyostelium discoideum*, associates transiently with cell contact areas, macropinosomes and phagosomes. *Mol. Biol. Cell*, **10**, 1205–1219.
11. Larochelle,D.A., Vithalani,K.K. and De Lozanne,A. (1997) Role of the *Dictyostelium* racE in cytokinesis: Mutational analysis and localization studies by use of green fluorescent protein. *Mol. Biol. Cell*, **8**, 935–944.
12. Seastone,D.J., Lee,E., Bush,J., Knecht,D. and Cardelli,J. (1998) Overexpression of a novel Rho family GTPase, RacC, induces unusual actin-based structures and positively affects phagocytosis in *Dictyostelium discoideum*. *Mol. Biol. Cell*, **9**, 2891–2904.
13. Chung,C.Y., Lee,S., Briscoe,C., Ellsworth,C. and Firtel,R.A. (2000) Role of Rac in controlling the actin cytoskeleton and chemotaxis in motile cells. *Proc. Natl Acad. Sci. USA*, **97**, 5225–5230.
14. Dumontier,M., Höcht,P., Mintert,U. and Faix,J. (2000) Rac1 GTPases control filopodia formation, cell motility, endocytosis, cytokinesis and development in *Dictyostelium*. *J. Cell Sci.*, **113**, 2253–2265.
15. Kay,R.R. and Williams,J.G. (1999) The *Dictyostelium* genome project: an invitation to species hopping. *Trends Genet.*, **15**, 294–297.
16. Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M., Yoshino,R., Mitra,B., Pi,M., Sato,T., Takemoto,K., Yasukawa,H., Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y. (1998) The *Dictyostelium* developmental cDNA project: generation and analysis of expressed sequence tags from the first-finger stage of development. *DNA Res.*, **5**, 1–7.
17. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
18. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **24**, 4876–4882.
19. Hall,T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Res. Symp. Ser.*, **41**, 95–98.
20. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
21. Page,R.D.M. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**, 357–358.
22. Newell,P.C., Telser,A. and Sussmann,M. (1969) Alternative developmental pathways determined by environmental conditions in the cellular slime mold *Dictyostelium discoideum*. *J. Bacteriol.*, **100**, 763–768.
23. Noegel,A.A., Metz,B.A. and Williams,K.L. (1985) Developmentally regulated transcription of *Dictyostelium discoideum* plasmid Ddp1. *EMBO J.*, **4**, 3797–3803.
24. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
25. Lohia,A. and Samuelson,J. (1996) Heterogeneity of *Entamoeba histolytica* rac genes encoding p21^{rac} homologues. *Gene*, **173**, 205–208.
26. Bourne,H.R., Sanders,D.A. and McCormick,F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117–126.
27. Wittinghofer,A. and Valencia,A. (1995) Three-dimensional structure of Ras and Ras-related proteins. In Zerial,M. (ed.), *Guidebook to the Small GTPases*. Oxford University Press, Oxford, UK, pp. 20–29.
28. Hirschberg,M., Stockley,R.W., Dodson,G. and Webb,M.R. (1997) The crystal structure of human rac1, a member of the rho family complexed with a GTP analogue. *Nature Struct. Biol.*, **4**, 147–152.
29. Ihara,K., Muraguchi,S., Kato,M., Shimizu,T., Shirakawa,M., Kuroda,S., Kaibuchi,K. and Hakoshima,T. (1998) Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue. *J. Biol. Chem.*, **273**, 9656–9666.
30. Hancock,J.F., Cadwallader,K., Paterson,H. and Marshall,C.J. (1991) A CAAX or CAAL motif and a second signal are sufficient for plasma membrane targeting of ras proteins. *EMBO J.*, **10**, 4033–4039.
31. Aravind,L. and Koonin,E.V. (1999) Fold prediction and evolutionary analysis of the POZ domain: structural and evolutionary relationship with the potassium channel tetramerization domain. *J. Mol. Biol.*, **285**, 1353–1361.
32. Ahmad,K.F., Engel,C.K. and Privé,G.G. (1998) Crystal structure of the BTB domain from PLZF. *Proc. Natl Acad. Sci. USA*, **95**, 12123–12128.
33. Feng,S., Cheng,J.K., Yt,H., Simon,J.A. and Schreiber,S.L. (1994) Two binding orientations for peptides of the Src SH3 domain: development of a general model for SH3–ligand interactions. *Science*, **266**, 1241–1247.
34. Nobes,C.D. and Hall,A. (1995) Rho, rac and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia and filopodia. *Cell*, **81**, 53–62.
35. Fraser,C.M. and Fleischmann,R.D. (1997) Strategies for whole microbial genome sequencing and analysis. *Electrophoresis*, **18**, 1207–1216.
36. Baldauf,S. and Doolittle,W.F. (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc. Natl Acad. Sci. USA*, **94**, 12007–12012.
37. Courjal,F., Chuchana,P., Theillet,C. and Fort,P. (1997) Structure and chromosomal assignment to 22q12 and 17qter of the ras-related Rac2 and Rac3 human genes. *Genomics*, **44**, 242–246.
38. Dietmaier,W. and Fabry,S. (1994) Analysis of the introns in genes encoding small G proteins. *Curr. Genet.*, **26**, 497–505.
39. Logsdon,J.M., Stolzfus,A. and Doolittle,W.F. (1998) Molecular evolution: Recent cases of spliceosomal intron gain? *Curr. Biol.*, **8**, R560–R563.