THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 16, Issue of April 18, pp. 10498–10505, 1997

Printed in U.S.A.

Tandem Duplication of *rab* Genes Followed by Sequence Divergence and Acquisition of Distinct Functions in *Trypanosoma brucei**

(Received for publication, October 29, 1996)

Helen Field and Mark C. Field‡

From the Laboratory of Cell Biology, Department of Biochemistry, Imperial College of Science, Technology, and Medicine, Exhibition Road, London SW7 2AY, United Kingdom

The Ras superfamily of small G proteins governs unidirectional cellular processes by virtue of GTP hydrolysis and concomitant conformational changes, which are in turn regulated by a number of accessory factors. Members of the Rab subfamily are important for correct targeting and fusion of intra-organellar vesicles loaded with trafficking proteins and lipids. During evolution from a prototype gene, novel functions may be acquired by duplicated daughter genes; for Rab proteins, this can be tested by location, which is specifically related to the function of each Rab. We have found an example of two *rab* genes in *Trypanosoma brucei* (*trab* genes) that clearly arose by tandem duplication, being highly related to each other and remaining juxtaposed in the genome, whose products have dramatically different subcellular locations, indicative of discrete functions. These two *trab* genes, isolated on a single genomic clone, are separated by a short intervening sequence and are in a head-to-tail orientation. The nucleotide sequences of the open reading frames and intervening sequence were determined and show that the genes are paralogues, probably arising from an ancient tandem duplication. Both genes are most homologous to *ypt1* and *sec4* in the *Saccharomyces cerevisiae* genome, while phylogenetic reconstruction indicates that although they have clearly diverged, the proteins are more closely related to each other than to other Rab protein sequences available in the data base. Immunofluorescence microscopy, using antibodies raised against the recombinant Trab proteins, clearly demonstrates that the native Trab proteins have completely distinct subcellular locations in the trypanosome. Trab1p is present in a widespread reticular location similar to BiP, suggesting an endoplasmic reticulum location, while Trab7p is observed in a discrete structure adjacent to the kinetoplast. Most interestingly, the Trab7p-positive compartment also appears to divide at the same time, or just prior to, the kinetoplast, *i.e.* early in mitosis, suggestive of association with structures in the flagellar pocket region. An estimate of the divergence time indicates that the *trab1*/*trab7* duplication occurred ~100 million years ago, and therefore, the persistence of this pair suggests an essential role in the survival of *T. brucei*.

Trypanosoma brucei spp., the causative agent of nagana in ungulates and sleeping sickness in humans, is a parasitic protozoan of economic and public health importance. Studies of this organism have revealed many unusual biological phenomena later observed in other systems, perhaps most remarkably *trans*-splicing (1) and glycosylphosphatidylinositol protein anchors (2). The paradigm glycosylphosphatidylinositol-anchored protein, the variant surface glycoprotein (VSG),[1] accounts for ~10% of total protein biosynthesis in the mammalian bloodstream form (BSF) of the parasite and provides an essential immunological barrier. The entire VSG coat, comprising 10^7 identical molecules, is replaced during antigenic switching in the BSF by a process of continued biosynthesis of new VSG and internalization/shedding of the old coat protein. Much VSG recycles: internalization is concomitant with surface delivery of newly synthesized VSG in the flagellar pocket (FP). In the insect procyclic form, VSG is replaced by the abundant glycosylphosphatidylinositol-anchored procyclic acidic repetitive protein (PARP) (3).

Rates of surface protein internalization are comparable with that of a mammalian cell, but in a trypanosome, only the FP, a small area of the cell surface, is available for exo- and endocytic activity due to a subplasmalemmal microtubular network over the rest of the cell body (4). The FP has four specialized adjacent microtubules running parallel down one side of the structure, coincident with a specialized area, the cytosome. This may be the site of exo- and endocytosis since it is associated with vesicles carrying VSG arising from the Golgi complex and with vesicular tubules and flattened cisternae structurally similar to the *trans*-Golgi network. Coated vesicles budding from the FP are observed only in BSF parasites (5). The FP is of interest for several reasons; first, for membrane trafficking of glycosylphosphatidylinositol-anchored glycoproteins; second, for sorting mechanisms, as endo- and exocytosis are presumably simultaneously active within a small membrane area (6, 7); and finally, as a potential site for therapeutic attack.

Detailed exploration of intracellular trafficking requires suitable markers, but the wide evolutionary distance between trypanosomes and most other eukaryotes is unfortunately reflected in a lack of immunological cross-reactivity between homologues. Marker proteins that are available include trypanosome BiP, an endoplasmic reticulum marker (8, 9), and CRAM, a cysteine-rich acidic transmembrane protein found in the FP and associated endocytic vesicles (10). Also available is an antibody to the CB1 epitope of a developmentally regulated 44-kDa protein of unknown function in the endocytic pathway,

* This work was supported by a project grant from the Wellcome Trust (to M. C. F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) X99951.

‡ To whom correspondence should be addressed. Tel.: 44-171-594-5277; Fax: 44-171-594-5207; E-mail: m.field@ic.ac.uk.

[1] The abbreviations used are: VSG, variant surface glycoprotein; BSF, bloodstream form; FP, flagellar pocket; PARP, procyclic acidic repetitive protein; PCR, polymerase chain reaction; EST, expressed sequence tag; ORF, open reading frame; kb, kilobase(s); IVS, intervening sequence (between ORFs); TLCK, N^{α}-*p*-tosyl-L-lysine chloromethyl ketone; PAGE, polyacrylamide gel electrophoresis; rTrab, recombinant Trab.

found in the FP, endosomes, and lysosomes (7).

The Rab family consists of GTPases with ~30% identity to Ras and with most homology confined to four GTP-binding motifs (11). Rab proteins are essential for protein trafficking in all eukaryotic organisms. Family members have low intrinsic GTPase activity, requiring activating proteins (GTPase-activating proteins) to increase turnover and act as effectors (12). GTP- and GDP-bound Ras have distinct conformations (13), and it can be inferred that Rab proteins will similarly undergo a conformational change upon GTP hydrolysis since their structure is highly homologous. Each Rab protein binds a subset of organelles; the C terminus is responsible for intracellular localization, while the N terminus recognizes the target organelle and is required for vesicle fusion (14). The paucity of markers suitable for trafficking studies led us to initiate the cloning of *rab* genes from *T. brucei* (15) since each Rab protein has a discrete subcellular location within the secretory pathway (16, 17). The study of *T. brucei* Rab proteins (Trab proteins) will yield new sequences important for targeting proteins to trypanosome organelles, provide information on the compartmentalization of trypanosomes, and generate subcellular markers.

A degenerate PCR cloning approach has yielded eight expressed sequence tags (EST) of small GTP-binding proteins from *T. brucei*, including Ran/TC4 (18) and seven Trab homologues (15), while separate studies have identified three additional *trab* genes (19, 20). We isolated two *trab* genes on a single genomic clone (15), and here we report the genomic organization of these genes (previously *rtb1* and *rtb7*) and the subcellular location of the encoded proteins.

MATERIALS AND METHODS

Nucleic Acids and Recombinant DNA Methods—Molecular biology-grade water was from an Elgastat Option 4 system (Elga) sterile-filtered through a 0.2-μm filter (Sartorius Corp.). Restriction and DNA-modifying enzymes were from Bioline, Promega, MBI Fermentas, New England Biolabs Inc., and Stratagene. Nitrocellulose was from Schleicher & Schuell (obtained from Anderman & Co. Ltd., London). Radioisotopes were from DuPont NEN and Amersham International. A P1 *T. brucei* library filter was a gift from Dr. S. Melville (Department of Pathology, Cambridge University, Cambridge, UK). Vector pGEX2tk and glutathione-Sepharose 4B were from Pharmacia Biotech Inc., and the λZAPII kit was from Stratagene. Molecular biology methods and antibody preparations followed standard procedures (21, 22). Plasmids were grown in *Escherichia coli* XL1-Blue (Stratagene) following transformation by electroporation with a BTX 600 ECM electroporator. PCR products and gel-embedded DNA were purified using PCR cleanup kits (Promega), and plasmid and λ DNAs were purified using QIAGEN kits following the manufacturers' instructions. Small-scale plasmid preparations were performed using the Promega Wizard system. Rabbit antibodies to *T. brucei* BiP were from J. Bangs (University of Wisconsin, Madison, WI). Secondary antibodies were from Sigma.

PCR—PCR was performed in 50-μl reactions with 25 pmol of each primer in ammonium buffer (Bioline), 2.5 mM $MgCl_2$, and 1 unit of *Taq* polymerase in a Model 480 thermal cycler (Perkin-Elmer) as follows: one cycle at 98 °C for 5 min; one cycle at 95 °C for 5 min; and 35 cycles at 94 °C for 1 min, 56 °C for 1 min, and 72 °C for 3.5 min. Templates were 1 μl of a 1:1000 dilution of R111 λ phage lysate or 1 μl of a λZAPII plaque equilibrated in 0.4 ml of SM phage dilution medium (21) or 10 ng of plasmid; phages were boiled for 10 min prior to PCR. The primers for determining orientation of *trab1* and *trab7* were RAT1FUS, RAT7PCR, and two reverse primers designed from the EST (15). The primers for subcloning the open reading frames (ORFs) into expression vectors were as follows: *trab7 5′*, GTGAAGTTTGGATGGATGAAAGAGGAACCC; *trab7 3′*, CCACCATAAGGAATTCAGCACCTCCACTT; *trab1 5′*, GAGTCCACTGGATCCATGATCACAGCAGC; and *trab1 3′*, CGACGAGAGGAATTCTTAACAGCAACCACCC. For expression of recombinant Trab proteins, each ORF was amplified from pBS.R7X1 (see below), digested with *Eco*RI/*Bam*HI for 16 h at room temperature, and ligated into *Eco*RI/*Bam*HI-cleaved pGEX2tk.

Southern Blotting—0.5-μg aliquots of R111 λ DNA were cut with the *Ava*II, *Bgl*II, *Cla*I, *Hin*dIII, *Hin*cII, *Pvu*II, *Sac*II, *Sal*I, *Xba*I, or *Xho*I restriction enzyme, and duplicate filters were prepared by high salt transfer Southern blotting (21). Filters were probed at high stringency (0.05 × SSC and 0.1% SDS at 75 °C) with *trab1* or *trab7* ESTs cut from pBluescript with *Eco*RI/*Bam*HI, gel-purified, and labeled using a random hexamer priming kit (Boehringer Mannheim). To map R111, 0.5-μg aliquots were cut with one or two restriction enzymes (*Ava*II, *Cla*I, *Sal*I, *Xba*I, and *Xho*I), blotted, and probed with the WD box primer (15) labeled with [γ-^{32}P]ATP using polynucleotide kinase at low and high stringency (6 × SSC and 0.1% SDS at room temperature, and 1 × SSC and 0.1% SDS at 65 °C, respectively). Single and double restriction digests of pBS.R7X1 were blotted and probed with *trab1* and *trab7* ESTs. To estimate the total number of *trab* genes in the *T. brucei* genome, a bacteriophage P1 library filter (SM5 No. 12) of *T. brucei* strain TREU 927/4 genomic DNA in grid format, with ~3 × haploid genome coverage, was probed with the WD primer at low stringency. Autoradiography was with X-Omat 5 film (Eastman Kodak Co.) at −85 °C with an intensifying screen; exposure times were 16 h to 7 days.

Construction of pBS.R7X1—A 6-kb *Xba*I fragment from R111 (15) was subcloned into pBluescript using a λZAPII kit; recombinant λ plaques (from the whole R111 *Xba*I digest) were screened on duplicate filters with *trab1* and *trab7* EST probes. Secondary screening was by PCR (15), and one clone was rescued to produce plasmid pBS.R7X1.

DNA Sequencing—Sanger sequencing of pBS.R7X1 through the ORFs and intervening region (IVS) was performed using a deaza-GTP kit (Pharmacia) and primer walking in both directions. Dye terminator cycle sequencing of untranslated regions and constructs was performed using a *Taq* thermal cycle sequencing kit and a Model 377 DNA sequencer (Perkin-Elmer) with polyethylene glycol-precipitated DNA (6.5% polyethylene glycol 6000 and 0.4 M NaCl on ice 20 min; centrifuged at 13,000 × *g* for 30 min at 4 °C; washed with 70% EtOH).

Antibodies—Expression plasmids pGEX2tk.Trab1 and pGEX2tk.Trab7 in *E. coli* were grown in L-broth and induced with 1.0 mM isopropyl-β-D-thiogalactopyranoside. Fusion proteins were affinity-purified on glutathione-Sepharose 4B and thrombin-cleaved to produce recombinant Trab1p and Trab7p, as described (23), with the following modifications. Protease inhibitors (pepstatin, leupeptin, antipain, and chymostatin, all at 2 μg/ml (protease inhibitor mixture)) and TLCK (dry) were added to the initial lysate. Cleared lysate was equilibrated with glutathione-Sepharose for 16 h at 4 °C. Thrombin treatment was at room temperature for 2 h, with more thrombin added after 1 h. Protein was quantitated by SDS-PAGE and Coomassie Blue staining. Typical yields were 0.4–4 mg of rTrab protein/liter of *E. coli*, with a purity of 95%. Rabbits were immunized three times with a total of 3 mg of rTrab1p or 0.9 mg of rTrab7p using the MPL® + TDM + CWS Adjuvant System (Sigma). Animals were exsanguinated, and serum was stored at 4 °C with azide. For affinity purification, 0.2 g of CNBr-Sepharose (Pharmacia) was activated and resuspended with an equal volume of coupling buffer (100 mM $NaHCO_3$ and 500 mM NaCl, pH 8.5) containing 0.5 mg of rTrab protein in phosphate-buffered saline and coupled overnight at 4 °C. Efficiency was ~45%. Purified antibodies were prepared from 50% ammonium sulfate-precipitated serum and stored at −85 °C.

Trypanosomes—Culture-adapted BSF *T. brucei* strain 427, a gift from S. Croft (London School of Hygiene and Tropical Medicine), was grown in HMI-18 (24) at 37 °C in a 5% CO_2 incubator. Procyclic form *T. brucei* strain 427 was grown in SDM79 (25); media and supplements were from Sigma. For boiling SDS lysates, freshly harvested trypanosomes were added to SDS-PAGE sample buffer (26) at 95 °C, heated for 5 min, and reduced with dithiothreitol. For hypotonic lysates, trypanosomes were resuspended in 10 mM Tris, pH 7.5, protease inhibitor mixture, phenylmethylsulfonyl fluoride, and TLCK, followed by two cycles of freeze/thaw and centrifugation at 13,000 × *g* for 15 min at 4 °C.

Western Blotting—Western blotting was performed on proteins electrophoresed on 17% SDS-polyacrylamide minigels and wet-blotted at 100 V for 3 h in transfer buffer (Hoefer Pharmacia Biotech Inc.) onto 0.45-μm nitrocellulose. Filters were blocked in 3 or 5% bovine serum albumin, phosphate-buffered saline, and 0.1% Tween 20 at 4 °C for 16 h. All remaining manipulations were at room temperature. Blots were probed with primary antibodies (in a 2–4-ml block) for 1 h, washed four times with phosphate-buffered saline/Tween 20, and then incubated with secondary antibody (horseradish peroxidase-conjugated goat anti-rabbit IgG diluted 1:2000) in blocking buffer for 20 min and washed. Detection was with either $NiCl_2$/diaminobenzidine (21) or ECL (Amersham Corp.) exposed to x-ray film. Purified antibodies at concentrations of 0.43 mg/ml (anti-Trab1p) and 0.17 mg/ml (anti-Trab7p) were typically used diluted by 1:1000 and 1:200, respectively.

Immunofluorescence Analysis—Immunofluorescence analysis was as described (8) with the following modification. Cells were fixed for 80 min with 1.5% formaldehyde and 0.32% glutaraldehyde (v/v). Purified

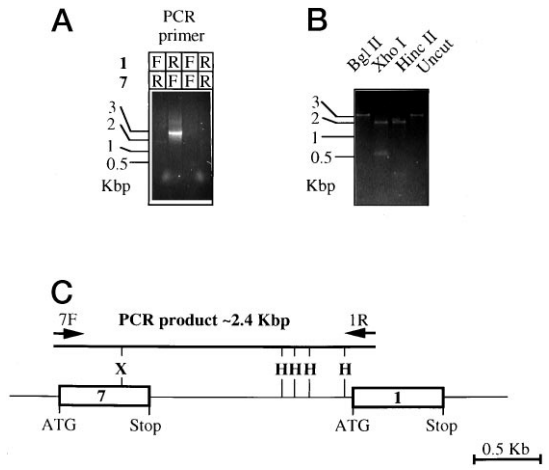

FIG. 1. ***trab7* precedes *trab1* in a head-to-tail orientation.** *A*, a PCR product was generated with only one combination of forward (*F*) and reverse (*R*) primers against the *trab1* and *trab7* ESTs. The PCR product was generated using genomic clone R111 as described under "Materials and Methods." *B*, the PCR product from *A* was cleaved by restriction enzymes known to separate the ESTs (by Southern blotting), *Xho*I and *Hinc*II, generating smaller fragments. *Bgl*II was used as a negative control. *C*, shown is a partial restriction map of the *trab7* and *trab1* locus, with sizes deduced from PCR data and restriction sites ascertained from sequence data. Genomic DNA is shown as a *thin line*, with the ORFs as *open boxes*. The PCR fragment in *A* is represented by the *boldface line*, with primers shown as *arrows*. *X*, *Xho*I restriction site; *H*, *Hinc*II restriction site.

anti-Trab antibodies were used at a 1:200 dilution. Fluorescence microscopy was performed on a Nikon Microphot-FX microscope using a 100× oil immersion objective and Ilford HP5 Plus ISO400 film.

Computer Analysis—The GenBank™ Data Bank was searched via the Internet using Netscape Version 2.0 (Netscape Communications Corp.) and FASTA or TXBLASTP against the non-redundant data base "nr" (National Center for Biological Information). The *Saccharomyces cerevisiae* genome was searched at the Stanford Medical School Web site.[2] Turbogopher was used for sequence retrieval from GenBank™. The University of Wisconsin Genetics Computer Group package was used at the Daresbury site (United Kingdom) via MacIP and Versaterm Pro for the following algorithms: SEQED, BESTFIT, GAP, PILEUP, MFOLD, SQUIGGLES (PLOTFOLD), COMPARE (DOTPLOT), and PEPTIDESTRUCTURE. Navigator and Autoassembler (Perkin-Elmer) and NIH Image were used on a Macintosh LCIII computer. For phylogenetic analysis, PAUP 3.1.1 (27) was used on a Macintosh Quadra 650; initial tree searches were done in branched and bound or exhaustive mode, with Ha-Ras assigned as outgroup. Preliminary results were bootstrapped (1000 replicates) in exhaustive mode. Divergence time was calculated using the program of Li (28). Three-dimensional computer modeling of Trab1p was performed automatically via SwissProt at EXPASY, and that of Trab7p was performed using INSIGHT II (Homology module) on an Indigo computer (Silicon Graphics Inc.) against the Ras pdb file (221P); four non-aligned loops were modeled *de novo* in free space and energy-minimized.

RESULTS

Organization of trab1 and trab7 Genes—Limited restriction endonuclease mapping of R111, a genomic clone containing both *trab1* and *trab7* (15), showed the insert was at least 26 kb long. The genomic organization of the *trab* genes was further explored by PCR using forward (F) and reverse (R) primers to each EST. We tested all four possible orientations: *trab1* and *trab7* oriented head-to-tail with *trab1* first (1F/7R), head-to-tail with *trab7* first (7F/1R), head-to-head (1F/7F), or tail-to-tail (1R/7R). A ~2.4-kb product was obtained using only the second primer combination, 7F/1R (Fig. 1*A*, *second lane*), indicating that *trab7* preceded *trab1* in a head-to-tail orientation, with an IVS of ~1.8 kb between ORFs. Identity of the PCR product was confirmed by digestion with *Xho*I and *Hinc*II (Fig. 1*B*), previously demonstrated to separate the ESTs in a Southern blot (Fig. 1*C*). The locations of the restriction sites were confirmed by sequence analysis (see below). The size of the IVS was initially overestimated due to anomalous migration of the PCR product; the gene organization was confirmed by sequence analysis (see below).

trab Genes in T. brucei—We analyzed clone R111 for additional *trab* genes to determine if *trab7* and *trab1* represent a simple duplication event or form part of a more extensive array. A degenerate oligonucleotide annealing to the conserved sequence WDTAGQE from the second GTP-binding site motif of Rab proteins (WD primer; see "Materials and Methods") was used in a low stringency Southern blot of genomic clone R111. No fragments other than those hybridizing to *trab1* and *trab7* probes were identified (data not shown), suggesting that these two *trab* genes do not form part of an extensive array. A 6-kb *Xba*I fragment from R111, which hybridized to *trab7* and *trab1* probes, was subcloned into pBluescript for sequence analysis (pBS.R7X1; see "Materials and Methods"). Mapping the position of the *trab7* gene demonstrated that R111 contains only ~2 kb of DNA upstream of the *trab7* ORF, and so the possibility of another *trab* gene lying upstream has not formally been excluded, although we believe it highly unlikely (see below and data not shown).

We also examined the total number of *trab* sequences in *T. brucei* by probing a ~3-fold degenerate (for haploid genome) *T. brucei* genomic library array with the WD box oligonucleotide at low and high stringency. Thirty-three positives were identified, corresponding to ~20 distinct *trab* loci. As previous Southern analysis and the studies of R111 have failed to identify extensive *trab* gene arrays (15), this number also provides an estimate of the total number of individual *trab* genes present in the *T. brucei* genome.

Sequence Data for trab1 and trab7 Genes and Their Intervening Sequence—We sequenced ~3.3 kb of subclone pBS.R7X1 including the *trab1* and *trab7* ORFs, the IVS, and the surrounding regions (Fig. 2). The *trab1* and *trab7* ORFs were identified by hypothetical translation of the nucleotide sequences. The orientation of the genes was confirmed, and the length of the IVS was determined to be 1165 base pairs. The ORF nucleotide sequences are 48.8% identical between *trab1* and *trab7*.

Both genes are expressed as assessed by Northern blotting and reverse transcription-PCR (15), and we identified consensus features for mRNA production (Fig. 2). Mini-exon splice sites and ATG start codons were located by comparing the sequence of the EST, derived from reverse transcription-PCR, with genomic sequences. A potential polyadenylation sequence for *trab7* was inferred from poly(T) tracts, with an appropriate GA addition site conforming to the consensus sequence (29). The predicted *trab7* message is 1.8 kb, consistent with Northern blotting data. The polyadenylation site of *trab7* is placed in very close juxtaposition to the mini-exon splice site of *trab1*: these two processes are known to be highly coordinated in trypanosomes (30). There is no polypyrimidine tract upstream of *trab7*, the first gene, suggesting that there is no gene upstream of *trab7* that is polyadenylated prior to the *trab7* splice site, further evidence that there are no other genes in this array.

trab1 and *trab7* messages are expressed at low levels compared with *T. brucei* Ran/TC4 mRNA (18) and other well characterized mRNAs such as PARP and VSG, all of which are high abundance messages. Also, these *trab* genes are constitutively expressed (15). As expected, none of the sequence motifs demonstrated to be involved in the maintenance of high expression levels (31) or stage-specific expression of PARP (32) or VSG (33)

[2] http://genome-www.stanford.edu.

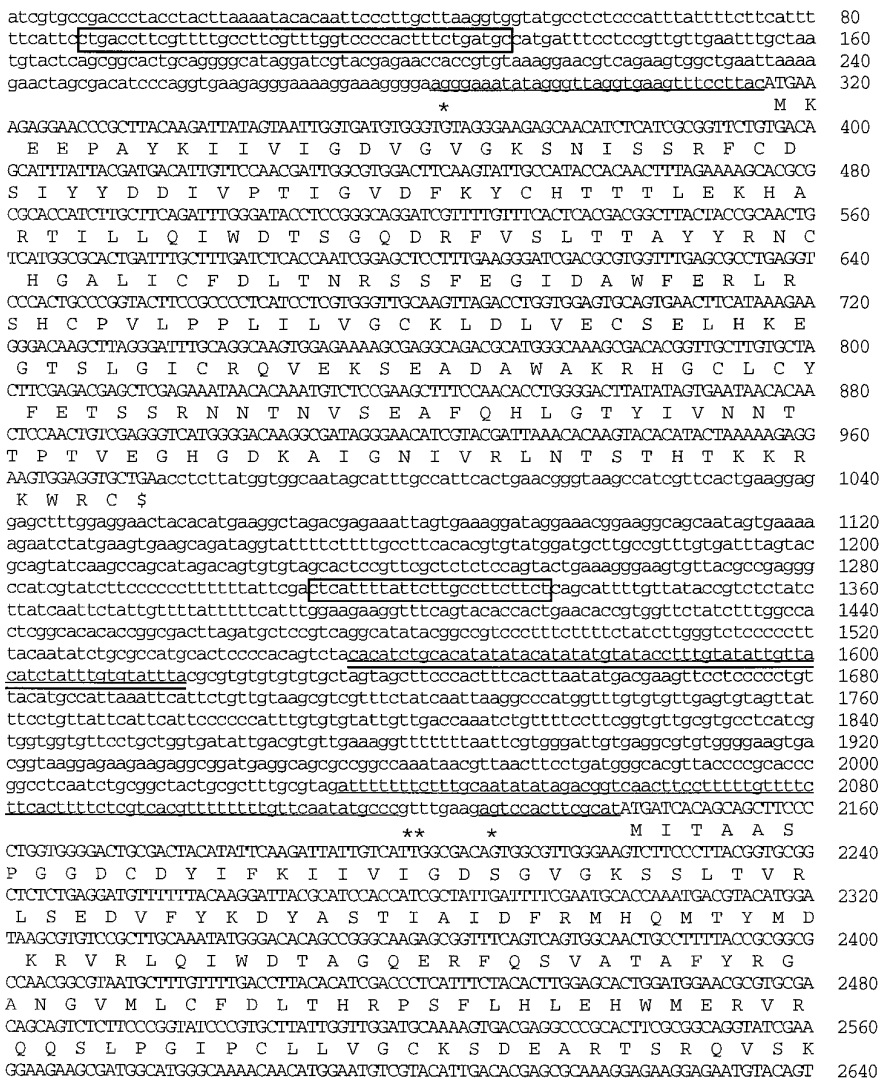

FIG. 2. **Sequence data confirm the orientation and close juxtaposition of the genes.** The nucleotide sequence of the genomic fragment containing both *trab* genes is shown. The *trab7* ORF precedes *trab1* (*upper-case*, ORFs; $, stop codon; *lower-case*, IVS/untranslated regions). Mini-exon splice sites and the putative polyadenylation site of *trab7* are *single-underlined*, with acceptor dinucleotides indicated by *asterisks* (*, mini-exon splice site; **, putative polyadenylation site). Three noncoding regions are homologous to one another (*boxed*) and were located using COMPARE. Regions with homology to the repeat units of the mini-exon donor gene (*double-underlined*; GenBank™ Data Bank accession number X00683) were located by BLAST search of an unidentified ORF found in the IVS against the nucleotide data base. Homology to this region in the *trab1* 3′-untranslated region was located using BESTFIT (70% nucleotide homology).

were present in the *trab1* and *trab7* genes. Although the structure of the *trab7* and *trab1* locus suggested a tandem duplication event, no strong candidates for possible recombination sites were found within the sequenced region of pBS.R7X1, *i.e.* related sequences upstream, downstream, and between *trab7* and *trab1*, nor were repetitive sequence motifs implicated in recombination between VSG genes observed (34). However, homology was found between the IVS (between genes) and a repetitive element in the *T. brucei* mini-exon donor gene, which in turn had homology to the reverse complement of a region in the 3′-untranslated region of *trab1* (*double-underlined* in Fig. 2), a potential Z-DNA site, and three other regions of internal homology were identified by the program COMPARE (*boxed* in Fig. 2). The presence of a limited stretch of repetitive DNA observed elsewhere in the *T. brucei* genome, may be significant for the evolution of this locus.

A divergence time of 100 million years ago was calculated from the nucleotide ORF sequences of *trab1* and *trab7* using the method of Li (28). This suggests that these genes are paralogues arising from an ancient duplication event. The homology of the amino acid sequence (64% similar) is greater than that of the nucleotide sequence (49%), so although synonymous nucleotide substitutions were saturated (28), considerable constraints on the divergence of the protein sequences are evident, indicating that the two gene products are important for *T. brucei* survival. The ORF sequences of *trab7* and *trab1* were used to search the data base, and both are clearly Rab homologues at the nucleotide and amino acid levels. However, Trab1p and Trab7p could not be unambiguously assigned as particular Rab homologues, which suggests a trypanosome-specific function for Trab1p and Trab7p. Because of species redundancy in the GenBank™ Data Bank, we also determined which genes in another unicellular eukaryote, *S. cerevisiae*, were most similar to *trab1* and *trab7*. *trab1* was most homolo-

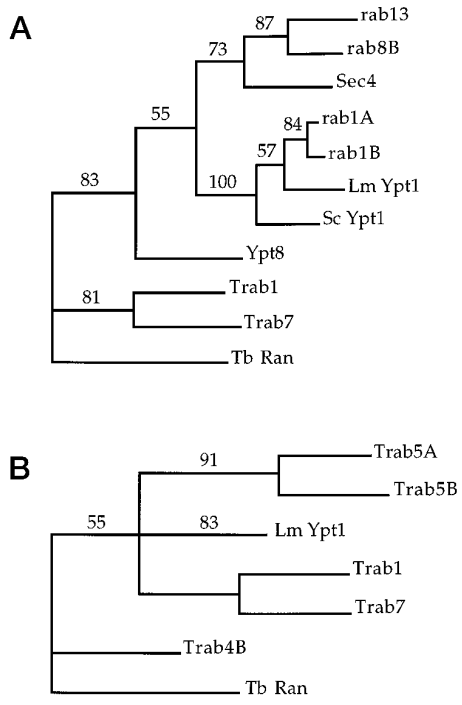

FIG. 3. **Trab1p and Trab7p are monophyletic.** *A*, the phylogeny of Trab1p and Trab7p hypothetical protein sequences was explored using PAUP in relation to a selection of protein homologues. Trab1p and Trab7p fall into a unique group, while all other Rab sequences group together, despite selection on the basis of their homology to both Trab1p and Trab7p. The tree was obtained from 1000 branched-and-bound bootstrap replicates after a branched-and-bound search: Rab1Ap and Rab13p, *Homo sapiens*; Sec4p, Sc Ypt1p, and Ypt8p, *S. cerevisiae*; Rab8Bp and Rab1Bp, *Rattus* spp.; and Lm Ypt1p (*L. major*). *B*, all kinetoplastid Ras superfamily protein sequences available were subjected to exhaustive phylogenetic reconstruction. *T. brucei* Rab5 sequences (Trab5Ap and Trab5Bp; see "Results") cosegregate, as do the Trab1p/Trab7p pair. *L. major* Ypt1p (*Lm Ypt1*) and Trab4Bp (*Tb*rab4B (19)) fall into neither family. The tree was constructed using exhaustive searching followed by bootstrapping with 1000 replicates. *T. brucei* Ran (*Tb Ran*) was chosen as an outgroup (18) for both trees, and numbers represent percent confidence for various branch points. Horizontal distances represent relative genetic distance, while vertical distances are for clarity only.

gous to *sec4*, while *trab7* was most similar to *ypt1*, which are themselves closely related. We analyzed the relationship of Trab1p and Trab7p to one another and to other homologous Rab sequences by phylogenetic reconstruction. Trab1p and Trab7p always cosegregated as a monophyletic group whether assessed against their closest Rab homologues (Fig. 3*A* and data not shown) or compared with kinetoplastid Rab sequences (Fig. 3*B*), including the putative Rab5 homologues Trab5Ap (*Tb*rab5 (19)) and Trab5Bp.[3]

Protein Sequences of Trab1p and Trab7p—Hypothetical translations of the Trab1p and Trab7p ORFs indicate that the protein sequences are 43% identical and 64% similar. Both sequences contain all of the major homology boxes typical of a Rab protein, including most of the residues involved in GTP/GDP binding and GTPase-activating protein binding (Fig. 4*A*). Trab1p also has a typical C-terminal amino acid motif for double geranylgeranylation, CCGG.

Interestingly, Trab7p contains several atypical features. Most significantly, the GTPase motif characterized by the WD box, normally WDTAGQE for Rab proteins (PM3 in Fig. 4*A*), is altered in Trab7p to WDTSGQD. The Glu → Asp change is not uncommon, but the Ala → Ser change (corresponding to a G → T nucleotide transversion in the first position of the codon) is highly unusual and possibly functionally significant; human RhoE also has an Ala → Ser substitution at this position and binds GTP but not GDP. Additionally, both Trab7p and RhoE have Ser at codon 12 (Ras numbering); combined with the Ala → Ser substitution, this results in lack of GTPase activity in RhoE (35). Also, constitutively active viral Ki-Ras has an Ala → Thr mutation in the WD box (36). Overall, these data suggest that Trab7p may possess abnormal GTPase activity.

Trab7p has a 12-amino acid insertion when compared with Ha-Ras, Ki-Ras, Ran, and at least 10 other Rab proteins (Fig. 4*A* and data not shown). Alignments place this insertion as an extension of loop 8 (nomenclature of Wittinghofer and Pai (13)). Residues important in the adjoining secondary structures are conserved in both Trab1p and Trab7p (Fig. 4*B*). The 12-residue loop of Trab7p was hypothetically large enough, by three-dimensional modeling, to interact with the adjacent nucleotide-binding site (data not shown).

Trab1p and Trab7p in Trypanosomes—To examine the expression and location of Trab1p and Trab7p, we raised polyclonal antisera to recombinant proteins. Affinity-purified anti-rTrab1p and anti-rTrab7p antibodies were used in Western analysis of trypanosome lysates. Specific signals were obtained from material migrating at 27 and 29 kDa, respectively, from both life stages (Fig. 5*A*). The theoretical molecular masses are 24 and 21.5 kDa for Trab1p and Trab7p, respectively, suggesting aberrant migration on SDS-PAGE. However, this was not due to post-translational modifications because the migration positions of rTrab1p and rTrab7p produced in *E. coli*, which does not isoprenylate proteins, were identical to those of the native trypanosome proteins (data not shown). Trab7p was also found to be highly labile and was seen only in fresh extracts.

Membrane Association of Trab Proteins—The unusual C-terminal sequence of Trab7p suggested that this protein may not be isoprenylated efficiently, and therefore, we wished to test if the protein was membrane-associated. Trab1p, by contrast, with its canonical C-terminal isoprenylation motif, is predicted to be isoprenylated and therefore membrane-associated. Procyclic cells were hypotonically lysed and centrifuged, and the soluble and particulate material was resolved by SDS-PAGE. Trab1p and Trab7p were located by Western blotting, and proteins were quantitated in identical samples (see "Materials and Methods"). Trypanosome proteins fractionated equally into soluble and particulate material, indicating efficient cell lysis, while both Trab1p and Trab7p appeared solely in particulate fractions (Fig. 5*B*). These data indicate that Trab7p is membrane-associated. In additional, our data are consistent with previous observations that some Rab proteins are exclusively membrane-bound and have no cytoplasmic pool (37).

Subcellular Locations of Trab1p and Trab7p—To begin to ascribe a function to Trab1p and Trab7p, we next chose to investigate the subcellular locations of the two proteins. We used affinity-purified antibodies to immunolocalize the native proteins by immunofluorescence analysis. Anti-Trab1p staining revealed a reticular pattern in both procyclic and BSF cells (Fig. 6, *A* and *C*) very similar to the appearance of cells stained for BiP, an endoplasmic reticulum luminal marker (Fig. 6, *B* and *D*). 10 μg of rTrab1p premixed with primary antibody for 10 min was sufficient to abolish the signal in BSF cells, demonstrating specificity (data not shown). Therefore, an endoplasmic reticulum location is tentatively assigned to Trab1p.

In contrast, cells stained with anti-Trab7p antibody revealed one or two discrete and highly compact structures in each trypanosome cell. The Trab7p-positive compartment was located close to the kinetoplast in both procyclic and BSF trypanosomes (Fig. 6, *E–J*). Prior incubation of the antibodies

[3] H. Field and M. C. Field, unpublished data.

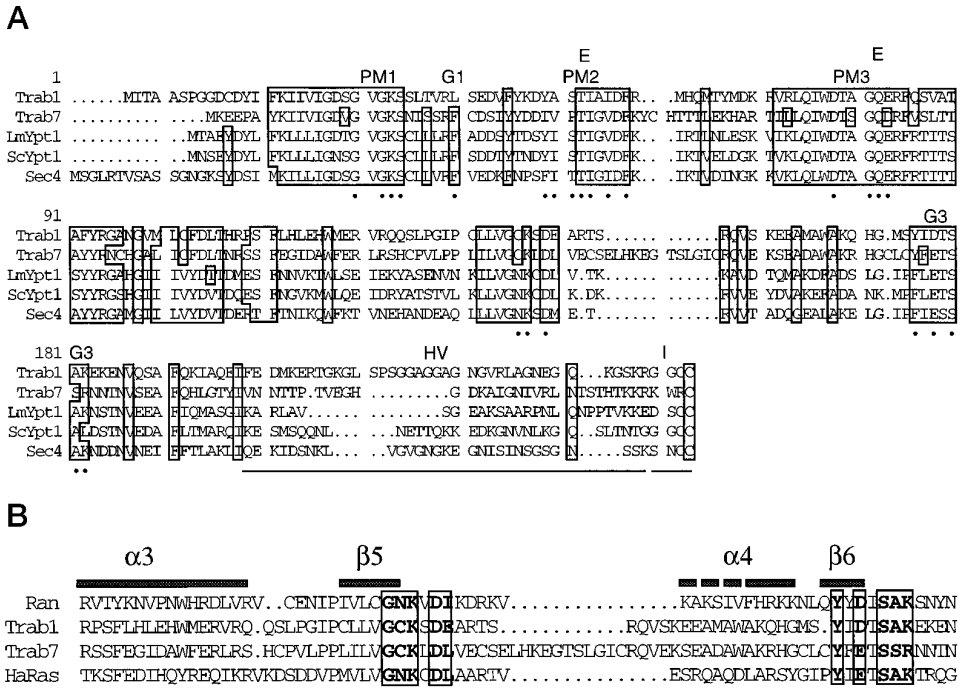

FIG. 4. **Trab1p and Trab7p are members of the Rab superfamily.** *A*, alignment of Trab1p, Trab7p, and a selection of Rab homologues. Conserved regions are *boxed* (46). Residues implicated in binding interactions, *i.e.* GTPase-activating protein binding (*E*), guanidine phosphate binding (*PM1–3*), and GTP/GDP binding (*G1–3*), are marked with *dots* under the sequence (see Ref. 11). The hypervariable region (*HV*) and the isoprenylation signal motif region (*I*) are *underlined*. *B*, alignment of the region containing the 10-amino acid loop from Trab7p (see *A*) together with Ha-Ras, human Ran, and Trab1p, showing secondary structure features derived from three-dimensional structures. α-Helical regions and β-sheets are indicated by *bars* (47). Residues vital for secondary structure (*boxed*) are conserved in both Trab1p and Trab7p.

with 10 μg of rTrab7p abolished the signal completely (data not shown). Most significantly, in cells containing two kinetoplast DNA networks (as revealed by Hoescht stain), two Trab7p-positive structures were observed positioned at either end of the kinetoplast (Fig. 6, *G–J*). The location of Trab7p is similar to that described for the basal bodies of trypanosomes (38), but clearly, further analysis is required to refine this assignment. Taken together, the immunolocalization data demonstrate radically different subcellular locations for Trab1p and Trab7p. Therefore, a tandem duplication resulted in acquisition of functional differences in these highly related Rab homologues.

DISCUSSION

We have identified a pair of small GTPase genes juxtaposed in a head-to-tail orientation in the genome of *T. brucei*. We were unable to find evidence for a reverse transcriptase-mediated duplication mechanism, *e.g.* mini-exon or poly(A) tail vestiges. While there are two genes present in the locus (*trab1* and *trab7*), five other *trab* genes and the gene for Ran/TC4 are dispersed and exhibit no linkage with each other (15, 18), suggesting that the *trab1*/*trab7* pair is unusual. In addition, all our data suggest that *trab1*/*trab7* are not part of a more extensive *trab* gene array. Phylogenetic reconstruction indicates that Trab1p and Trab7p are more similar to each other than to any other Rab sequence, evidence that they are paralogues. Taken together, these data strongly argue that these genes are the result of an ancient tandem duplication event, estimated to have occurred ~100 million years ago.

Duplication, probably by homologous recombination, has resulted in many gene arrays in *T. brucei*, *e.g.* the PARP locus with arrays of two or three genes (31) and the extensive tubulin locus (39). Homologous recombination is highly active (40–42) and is probably responsible for producing and manipulating the reservoir of up to 1000 different VSG genes providing diverse antigenicity (34, 43). Tandem arrangement of genes allowing polycistronic expression is a common strategy in *T. brucei* and may facilitate coordination of gene expression. Indeed, *trab1* and *trab7* are transcribed at low levels and in equal amounts (15).

By probing of a genomic library of known complexity, the total number of Trab-related genes was estimated to be ~20. Since 11 *trab* clones and ESTs have been reported to date, it is likely that a representative sampling of *trab* sequences, covering most of the trypanosome intracellular trafficking pathways, is already in hand. This number is similar to that obtained for *S. cerevisiae*, also a unicellular eukaryote, and is in agreement with our earlier estimate of total *trab* numbers (15). Therefore, these data suggest that, as a first approximation, *T. brucei* possesses a secretory pathway of equivalent complexity to yeast.

We propose that the continued linkage of *trab1* and *trab7* is due to functional selection. Conservation of protein sequence contrasted with saturated synonymous nucleotide replacement strongly suggests a functional selection on the gene products. While both Trab sequences are homologous to a selection of Rab sequences in the data base, neither exhibited strong homology to a particular Rab sequence, which suggests instead that they are trypanosome-specific. The immunolocalization data for Trab7p, indicating association with the kinetoplast and FP, are clearly consistent with this proposal. In addition, we were unable to detect Trab1p or Trab7p in COS-7 cells, either by Western blotting or by immunofluorescence analysis, or in *Leishmania major*, by Western blotting.[3] While the former result was not unexpected, the lack of immunological cross-reactivity between Trab proteins and *Leishmania* Rab proteins is also consistent with a species-specific function.

Trab1p has canonical Rab sequences, including a C-terminal isoprenylation signal, and shows diffuse reticular staining consistent with a Ypt1p/Rab1p function. The homology to *sec4* (and the closely related *ypt1*) detected by searching the *S. cerevisiae* genome may reflect this. In contrast, Trab7p has several un-

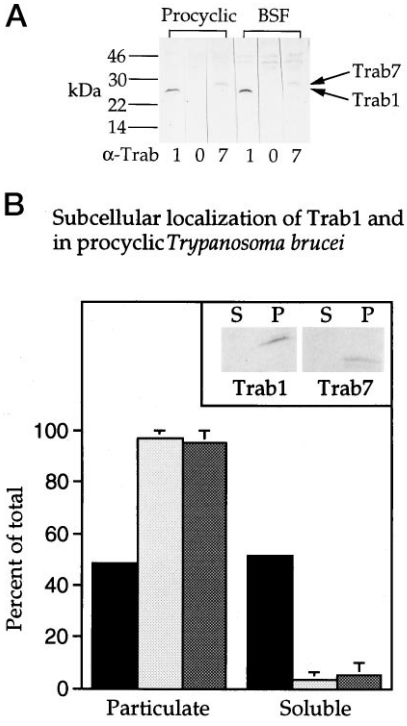

FIG. 5. **Antibody detection of Trab proteins and their membrane localization.** *A*, affinity-purified antibodies raised against rTrab1p and rTrab7p (*lanes 1* and *7*, respectively) detected specific antigens in *T. brucei* lysates from both procyclic and BSF trypanosomes in Western blotting. No cross-reactivity was observed between antibodies to Trab1p and Trab7p. Primary antibody was omitted from the negative controls (*lanes 0*). Migration positions of molecular mass standards (in kDa) are indicated to the left. *B*, Trab1p and Trab7p were detected by Western blotting in procyclic trypanosome lysates following separation into particulate (*P*) and soluble (*S*) fractions (*inset*). Antigens reactive with anti-Trab1p and anti-Trab7p antibodies were detected only in the insoluble particulate fraction. The percentage of each antigen detected in particulate and soluble fractions (*main panel*) was quantitated by scanning the developed Western blot using NIH Image. Trypanosome lysis was efficient, as judged by the release of proteins to the soluble fraction, quantitated by scanning a Coomassie Blue-stained gel (*main panel*, ■). Rabbit antisera raised against rTrab1p (▨) or rTrab7p (▨) were affinity-purified and used in Western blots against whole trypanosome lysates or fractionated proteins from 10^7 cells (see "Methods and Materials"). Detection was by $NiCl_2$/diaminobenzidine.

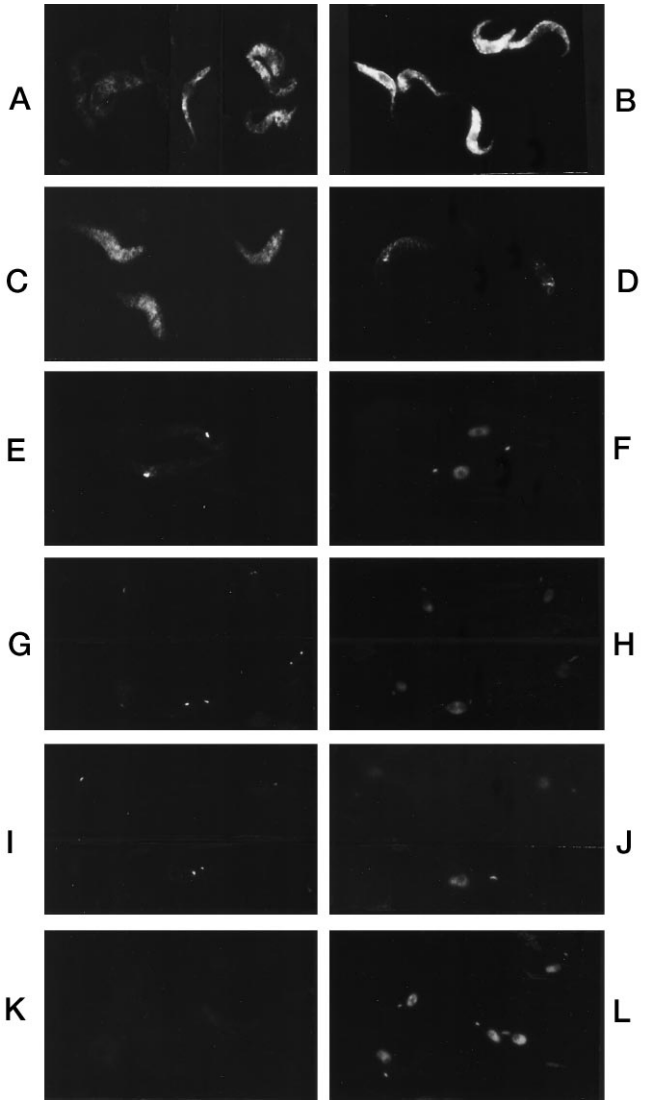

FIG. 6. **Subcellular locations of Trab1p and Trab7p in *T. brucei*.** Native Trab proteins were detected in whole trypanosome cells by indirect immunofluorescence using affinity-purified antibodies specific for each Trab protein. Procyclic (*A*, *B*, *E–H*, *K*, and *L*) or BSF (*C*, *D*, *I*, and *J*) cells were stained with anti-Trab1p antibody (*A* and *C*), anti-*T. brucei* BiP antibody (*B* and *D*), anti-Trab7p antibody (*E*, *G*, and *I*), or Hoescht stain (*F*, *H*, *J*, and *L*) of the same fields as *E*, *G*, *I*, *K*, respectively. Anti-Trab1p staining in procyclic cells gives reticular staining (*A*) comparable to that given by anti-*T. brucei* Bip antibody (*B*). Identical staining patterns are seen in BSF cells (*C* and *D*). Note that the exposure of *D* is lower than that of *B*; staining intensities were actually equivalent. Anti-Trab7p antibodies detect a discrete compartment close to the kinetoplast in procyclic cells, and two Trab7p entities are seen in cells with two kinetoplasts (*E–H*). Trab7p entities are also found in BSF cells (*I* and *J*). Primary antibody was omitted from the negative controls, shown at the same exposure as stained cells (*K*).

usual features. An important alteration in Trab7p is an Ala → Ser substitution in the WD box, potentially resulting in a GTP-bound, constitutively active protein like RhoE (35) or viral Ki-Ras, which has an Ala → Thr mutation (36). A second change in Trab7p is a 12-amino acid loop adjacent to the nucleotide-binding pocket. We believe it is likely that Trab7p binds only GTP and does not hydrolyze it, but clearly, this needs to be shown empirically. The subcellular location of Trab7p is also highly unusual, consisting of a single spot in nondividing cells located close to the kinetoplast. This structure divides at a similar time to the kinetoplast, clearly preceding nuclear division, suggesting association of the Trab7p compartment with basal body/FP structures. Like the Trab7p compartment, the basal body and FP are located close to the kinetoplast and divide with it, prior to nuclear division (38).

Carboxyl-terminal isoprenylation is essential for Rab function and occurs in *T. brucei* (44), and in all Rab homologues reported in *T. brucei* with the exception of Trab7p (KWRC), canonical isoprenylation sequences are found, *i.e.* SCAC (Trab4Bp), GCAC (Trab5Ap), GCCG (Trab5Bp), CCGG (Trab1p), and DSCC (Ypt1p from kinetoplastid *L. major*). Despite its unusual signal sequence, native Trab7p is membrane-associated, which suggests that it is a substrate for the trypanosome isoprene transferases. It remains to be determined which isoprene is present on Trab7p.

The divergence of Trab7p from the canonical Rab sequence and the association with a kinetoplast-specific organelle are suggestive of a novel role for Trab7p in the trypanosome cell. We propose that *trab7* arose by duplication of a progenitor *trab* gene, with Trab1p now fulfilling a more basic endoplasmic reticulum-associated function and Trab7p co-opted into a specialized role. Trab1p and Trab7p are 43% identical and 64% similar, reminiscent of Arf1 and Arf6, which are 66% homologous and also have different locations (45). Because of the absence of clear homologues for Trab1p and Trab7p, the precise functions of these proteins must be determined directly, which

is being actively pursued. Efforts to further localize the individual proteins and to gain more insight into their function by overexpression and direct study of the GTPase activities of these potentially important trypanosome proteins are currently under way. In summary, we report the presence of a tandem pair of *rab* genes in the protozoan parasite *T. brucei*. Immunolocalization and sequence data suggest a species-specific role for at least one (Trab7p). The *trab1*/*trab7* pair has remained together for a considerable period of time, indicative of functional importance.

Acknowledgments—We thank Anneke Loos for sequencing IVS and mapping studies, Jay Bangs for antibodies to trypanosome BiP, Simon Croft for culture-adapted BSF trypanosomes, Richard Thomas (Museum of Natural History, London) for analysis of the divergence time, and Ann Ridley for the thrombin cleavage protocol.

REFERENCES

1. Agabian, N. (1990) *Cell* **61,** 1157–1160
2. Cross, G. A. M. (1990) *Annu. Rev. Cell Biol.* **6,** 1–39
3. Mowatt, M. R., and Clayton, C. E. (1987) *Mol. Cell. Biol.* **7,** 2838–2844
4. Duszenko, M., and Seyfang, A. (1993) in *Advances in Cell and Molecular Biology of Membranes* (Tartakoff, A. M., and Plattner, H., eds) pp. 227–258, Jai Press Inc., Greenwich, CT
5. Clayton, C., Hausler, T., and Blattner, J. (1995) *Microbiol. Rev.* **59,** 325–344
6. Webster, P., and Russell, D. G. (1993) *Parasitol. Today* **9,** 201–206
7. Brickman, M. J., and Balber, A. E. (1994) *J. Eukaryotic Microbiol.* **41,** 533–536
8. Bangs, J. D., Uyetake, L., Brickman, M. J., Balber, A. E., and Boothroyd, J. C. (1993) *J. Cell Sci.* **105,** 1101–1113
9. Bangs, J. D., Brouch, E. M., Ransom, D. M., and Roggy, J. L. (1996) *J. Biol. Chem.* **271,** 18387–18393
10. Lee, M. G.-S., Bilhain, B. E., Russell, D. G., Deckelbaum, R. J., and Van der Ploeg, L. H. (1990) *Mol. Cell. Biol.* **10,** 4506–4517
11. Zerial, M., and Huber, L. A. (eds) (1995) *Guidebook to the Small GTPases*, pp. 12–29, Oxford University Press, Oxford
12. Novick, P., and Brennwald, P. (1993) *Cell* **75,** 597–601
13. Wittinghofer, A., and Pai, E. (1991) *Trends Biochem. Sci.* **16,** 383–387
14. Steele-Mortimer, O., Clague, M. J., Huber, L. A., Chavrier, P., Gruenberg, J., and Gorvel, J.-P. (1994) *EMBO J.* **13,** 34–41
15. Field, M. C., and Boothroyd, J. C. (1995) *Exp. Parasitol.* **81,** 313–320
16. Pfeffer, S. R. (1992) *Trends Cell Biol.* **2,** 41–46
17. Nuoffer, C., and Balch, W. E. (1994) *Annu. Rev. Biochem.* **63,** 949–990
18. Field, M. C., Field, H., and Boothroyd, J. C. (1995) *Mol. Biochem. Parasitol.* **69,** 131–134
19. El-Sayed, N. M. A., Alarcon, C. M., Beck, J. C., Sheffield, V. C., and Donelson, J. E. (1995) *Mol. Biochem. Parasitol.* **73,** 75–90
20. Klöckner, T., van den Bogaard, M., and Boshart, M. (1995) *Molecular Parasitology Meeting, September 17–21* (Goldberg, D., Parsons, M., and Ullu, E., eds) p. 123, Marine Biological Laboratory, Woods Hole, MA
21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
22. Harlow, E., and Lane, D. (eds) (1988) *Antibodies: A Laboratory Manual*, pp. 53–553, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
23. Ridley, A. J., and Hall, A. (1992) *Cell* **70,** 389–399
24. Hirumi, H., and Hirumi, K. (1989) *J. Parasitol.* **75,** 985–989
25. Cunningham, I. (1977) *J. Protozool.* **24,** 325–329
26. Laemmli, U. K. (1970) *Nature* **227,** 680–685
27. Swofford, D. L. (1993) *Phylogenetic Analysis Using Parsimony*, Version 3.1.1, Illinois Natural History Survey, Champaign, IL
28. Li, W.-H. (1993) *J. Mol. Evol.* **36,** 96–99
29. Field, H., Cerami, A., and Henderson, G. B. (1992) *Mol. Biochem. Parasitol.* **50,** 47–56
30. Graham, S. V. (1995) *Parasitol. Today* **11,** 217–223
31. Clayton, C., and Hotz, H.-R. (1996) *Mol. Biochem. Parasitol.* **77,** 1–6
32. Hehl, A., and Roditi, I. (1994) *Parasitol. Today* **10,** 442–445
33. Berberof, M., Vanhamme, L., Tebabi, P., Pays, A., Jefferies, D., Welburn, S., and Pays, E. (1995) *EMBO J.* **14,** 2925–2934
34. Field, M. C., and Boothroyd, J. C. (1996) *J. Mol. Evol.* **42,** 500–511
35. Foster, R., Hu, K.-Q., Lu, Y., Nolan, K. M., Thissen, J., and Settleman, J. (1996) *Mol. Cell. Biol.* **16,** 2689–2699
36. Shimizu, K., Birnbaum, D., Ruley, M. A., Fasano, O., Suard, Y., Edlund, L., Taparowsky, E., Goldfarb, M., and Wigler, M. (1993) *Nature* **304,** 497–500
37. Zerial, M., Parton, R., Chavrier, P., and Frank, R. (1992) *Methods Enzymol.* **219,** 398–409
38. Robinson, D. R., and Gull, K. (1991) *Nature* **352,** 731–733
39. Thomasshow, L. S., Milhausen, M., Rutter, W. J., and Agabian, N. (1983) *Cell* **32,** 35–43
40. ten Asbroek, A. L. M. A., Ouellette, M., and Borst, P. (1996) *Nature* **348,** 174–175
41. Eid, J., and Sollner-Webb, B. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 2118–2121
42. Li, F., and Gottesdiener, K. M. (1996) *Nucleic Acids Res.* **24,** 534–535
43. Van der Ploeg, L. H. T., Valerio, D., de Lange, T., Bernards, A., Borst, P., and Grosveld, F. G. (1982) *Nucleic Acids Res.* **10,** 5905–5923
44. Field, H., Blench, I., Croft, S., and Field, M. C. (1996) *Mol. Biochem. Parasitol.* **82,** 67–80
45. Peters, P. J., Hsu, V. W., Ooi, C. E., Finazzi, D., Teal, S. B., Oorschot, V., Donaldson, J. G., and Klausner, R. D. (1995) *J. Cell. Biol.* **128,** 1003–1017
46. Zahraoui, A., Touchot, N., Chardin, P., and Tavitian, A. (1989) *J. Biol. Chem.* **264,** 12394–12401
47. Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W., and Wittinghofer, A. (1995) *Nature* **374,** 378–381