J Mol Evol (1997) 44:159–168

JOURNAL OF MOLECULAR EVOLUTION

Isolation and Characterization of Five Actin cDNAs from the Cestode *Diphyllobothrium dendriticum:* A Phylogenetic Study of the Multigene Family

Monica H. Wahlberg,[1] Mark S. Johnson[2]

[1] Department of Biology, Åbo Akademi University, Artillerigatan 6, FIN-20520 Åbo/Turku, Finland
[2] Molecular Modelling and Biocomputing Group, Turku Centre for Biotechnology, University of Turku and Åbo Akademi University, Tykistökatu 6, FIN-20521 Turku, Finland

Received: 19 June 1996 / Accepted: 20 August 1996

Abstract. Five cDNAs (*pDidact2–pDidact6*), representing different actin genes, were isolated from a *Diphyllobothrium dendriticum* cDNA library, and the DNA as well as the putative amino acid sequences were determined. The corresponding *Didact2* and *Didact4* genes code for peptides 376 amino acids long, with molecular weights 41,772 and 41,744 Da, respectively, while the deduced *Didact3* protein is 377 amino acids long and weighs 41,912 Da. The *pDidact5* and *-6* cDNAs lack nucleotides corresponding to three to six amino acids at the amino-terminus. Two of the five cDNAs contain the conventional AATAAA as the putative polyadenylation signal, one has the common variant ATTAAA, whereas the hexanucleotide AATAGA is found 15 and 18 nucleotides, respectively, upstream of the poly(A) site in two of the cDNAs. Phylogenetic studies including 102 actin protein sequences revealed that there are at least four different types of cestode actins. In this study three of these types were found to be expressed in the adult *D. dendriticum* tapeworm. Structurally the cestode actin groupings differ from each other to an extent seen only among the metazoan actins between the vertebrate muscle and cytoplasmic isoforms. In the phylogenetic trees constructed, cestode actins were seen to map to two different regions, one on the border of the metazoan actins and the other within this group. It is, however, difficult to say whether the cestode actins branched off early in the metazoan evolution or if this position in the phylogenetic tree only reflects upon differences in evolutionary rate.

Key words: *Diphyllobothrium dendriticum* — Cestode — Actin cDNA — Polyadenylation signal — Multigene family — Molecular evolution — Phylogeny

Introduction

Actins are highly conserved ubiquitous proteins which generally are encoded by multigene families. Such families are thought to have arisen by an initial gene duplication and divergence of a common ancestral gene (Maeda and Smithies 1986). Several studies have been made where the phylogenetic relationships between actin protein sequences are investigated. These include studies where a wide range of actin proteins have been analyzed (Hightower and Meagher 1986; Meagher 1991; Reece et al. 1992; Mounier et al. 1992; Kovilur et al. 1993; Hennessey et al. 1993; Sheterline et al. 1995; Drouin et al. 1995) and investigations where actins from only a few specific species or phyla have been dealt with (Miwa et al. 1991; Fang and Brandhorst 1994). Due to the high sequence conservation among most actins, phylogenetic analyses produce few groupings which can be considered significant (Hennessey et al. 1993; Sheterline et al. 1995; Mounier et al. 1992). Thus, the evolutionary history of the actins is a source of continuous study and controversy.

On the basis of both the primary protein sequence and

Correspondence to: M.H. Wahlberg

tissue localization, the actins of vertebrates and of the arthropods *Bombyx mori* and *Drosophila melanogaster* are classified into two main types: the muscle and the nonmuscle actins (Vandekerckhove and Weber 1978; Mounier et al. 1992). The arthropod muscle actins, however, differ from the vertebrate muscle isoforms to such an extent that it seems likely that muscle actin genes have arisen independently both within the chordate and the arthropod phyla. This suggests that muscle actins arose from cytoplasmic isoforms at least twice during animal evolution (Mounier et al. 1992). The urochordate muscle actins possibly represent a transition from a non–muscle-like sequence to a vertebrate muscle-like sequence. Of 21 diagnostic muscle-specific amino acids in vertebrate cardiac muscle actin, five or six are differing when compared with *Styela* (urochordate) muscle actins, while 15 different amino acids are found in comparison with muscle actin from the sea urchin (Kovilur et al. 1993). Generally, however, the invertebrate actins are considered to be of the vertebrate cytoplasmic actin type.

In this study we have investigated the actin multigene family of the seagull tapeworm *Diphyllobothrium dendriticum.* The flatworms are thought to have separated early in metazoan evolution (Hori and Osawa 1987; Field et al. 1988; Adoutte and Philippe 1993) and it has been suggested that ancestral flatworm-like animals would occupy a position as the likely ancestors of most metazoans (Barnes et al. 1988). Therefore, studies on gene families in the flatworms living today might provide important clues to gene and organismal evolution.

When screening a *D. dendriticum* cDNA library, six different actin genes expressed in the adult tapeworm were found (one is published in Wahlberg et al. 1994). These were characterized and could, on the basis of their sequences be divided into three clearly different groups. The cestodes *Taenia soilum* and *Echinococcus granulosus* also possess actin proteins belonging to these groups. The *E. granulosus* act2_echgr protein, however, was found to cluster by itself.

Materials and Methods

Parasite Materials. Golden hamsters were experimentally infected with plerocercoids of *D. dendriticum* obtained from whitefish (*Coregonus lavaretus*) from Lake Pyhäjärvi in southwest Finland. After 10–14 days adult tapeworms were collected by flushing the intestine of infected hamsters with 0.9% NaCl.

RNA Isolation and Construction of a cDNA Library. RNA was isolated from adult *D. dendriticum* by the acid-guanidinium-phenol-chloroform method, essentially as described by Chomczynski and Sacchi (1987) and a Uni-ZAP XR cDNA library (Stratagene) was constructed.

Screening the cDNA Library. Some 200,000 recombinant phages of the cDNA library were screened according to Benton and Davis (1977). Prehybridization and hybridization were carried out at 42°C in 30% formamide, 5× SSC (1× SSC is 150 mM NaCl, 15 mM sodium citrate), 5× Denhardt's solution (0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.1% bovine serum albumin), 0.02 M sodium phosphate pH 6.5, 0.2% SDS (sodium dodecyl sulphate), and 100 $\mu g\ ml^{-1}$ tRNA for 4 and 24 h, respectively. A fragment of 328 bp spanning from nucleotide 520 to 834 in the *Didact1* actin cDNA from *D. dendriticum* (Wahlberg et al. 1994), including 14 bases from the pBluescript vector, was random primed labeled with [α-^{32}P]dCTP and used as a probe. The washing conditions were 2× SSC at 42°C for 20 min. The plaques were purified to homogeneity by consecutive platings and rescreenings.

Sequence Analysis. In vivo excisions and plasmid preparations were carried out according to protocols from Stratagene. DNA deletions made with the ExoIII/Mung Bean kit (Stratagene) were subcloned into pBluescript SK(−) (Stratagene). Nucleotide sequences were determined by the dideoxy chain termination procedure of Sanger et al. (1977) using the ^{T7}Sequencing kit from Pharmacia Biotech and [α-^{35}S]dATP. After control sequencing, computer analysis of sequence data was performed with the Sequence Analysis Software Package program, version 8.1 (1995), from the Genetics Computer Group of the University of Wisconsin. Actin protein sequences for comparison studies were taken from the SwissProt database release 31.0 (3/95) and nucleic acid sequences from the GenBank database release 89.0 (6/95). Sequence comparisons were made with the Gap program and molecular weight calculations with Peptidesort.

Phylogenetic Studies. Given the uncertainties in trees derived for actins from sequence analysis (Mounier et al. 1992; Hennessey et al. 1993; Sheterline et al. 1995), we chose to use an alternative strategy to make the best use of our available data: (1) To compare actins with a sequence comparison matrix derived from observed amino acid replacements from sets of protein structures from the same percentage identity range covered by the actin sequences themselves. (2) To use multivariate analysis to define clusters of similar actin sequences. (3) To merge the members of these clusters into a single group, thus representing each group by the average of the original distances of its members to all other actins. Phylogenetic trees were then derived from these average data. (4) To treat the more similar main metazoan actin group separately but employing the same strategy as outlined in 2 and 3 above.

All full-length actin amino acid sequences from the SwissProt database 31.0 (3/95) were obtained, as well as five amino acid sequences corresponding to the six *D. dendriticum* actin cDNAs from this and an earlier study (Wahlberg et al. 1994). Two of the *D. dendriticum* nucleic acid sequences, *Didact1* and *-4*, correspond to identical proteins. (The sequences used are listed at the end of this section.) The 102 sequences were aligned with the computer program MALIGN (Johnson et al. 1993) and a structure-based scoring matrix (Johnson and Overington 1993) derived from pairs of aligned protein structures with percentage identities >80% (obtainable from http://www.btk.utu.fi).

Pairwise distances (D) were obtained from the multiple alignments: $D = -100$ ln NAS, where the alignment score (summed scores for matching aligned residues less penalties assessed for indels in the alignment) is normalized for the length of the shorter of the two sequences and the maximum possible alignment score for the pair (NAS). A constant penalty of 40 for insertions/deletions was used.

Principal components analysis (PCA) is a standard multivariate statistical analysis procedure (Chatfield and Collins 1989) that can be used to establish clusters of data with similar features. In this case, the input to the program consists of the matrix of distances (D above) obtained from each pair of aligned sequences. The output of the program consists of coordinates for each sequence in which the overall variance is a maximum. Since PCA is a multidimensional approach, these coordinates can be output in two, three, four, or more dimensions (up to 102 dimensions, although the information content provided by most dimensions will be nil). For our purposes, sequences were clustered based on the Euclidean distances calculated between the coordi-

nates of sequence pairs in the 11 most informative dimensions. In considering all 102 sequences, those within a Euclidean distance of 0.65 of each other defined a cluster (maximum distance 1.83); for the 40 sequences of the main metazoan group, an empirically chosen cutoff of 0.45 was used (maximum distance 1.88).

The original distances obtained from the sequence alignments were then averaged to provide single values representing the relationship of one cluster to another cluster. Although this results in a loss of details about individual clusterings (which can be reconstructed later), fluctuations within the family, due to an extra amino acid replacement or two, can be minimized. Trees were constructed using programs of the Phylip package (SEQBOOT, NEIGHBOR, FITCH, CONSENSE, DRAWGRAM) of Felsenstein (1985). Group averaged distances were calculated from 1,000 individual trees (global rearrangement, best tree from random orders of the input data each).

The actin amino acid sequences used in the phylogenetic studies are listed below. The Roman numerals indicate the actin groupings and the Arabic numerals refer to the number codes in the 2D projection of Fig. 3. Individual sequences that in the phylogenetic trees of Figs. 4 and 5 cluster together with a defined group, are in this list included as parts of that particular group (act_volca is included in the planta actin group and actd_phypo with the protista-II sequences).

PLATYHELMINTHES (CESTODA): 1. *Didact1* and *-4 Diphyllobothrium dendriticum* +act_taeso *Taenia solium* (I), 2. *Didact2 D. dendriticum* (I), 30. *Didact3 D. dendriticum* (II), 31. *Didact5 D. dendriticum* (II), 60. act1_echgr *Echinococcus granulosus* (III), 61. *Didact6 D. dendriticum* (III), 62. act2_echgr *E. granulosus* (IV); VERTEBRATA, CYTOPLASMIC ISOFORMS (c): 3. actb_human *Homo sapiens,* 4. actb_rabit *Oryctolagus cuniculus* (rabbit), 5. actb_cypca *Cyprinus carpio* (common carp), 6. actg_human *H. sapiens,* 7. actb_ansan *Anser anser anser* (western graylag goose), 8. actb_xenbo *Xenopus borealis* (Kenyan clawed frog); TUNICATA (c): 46. actc_stypl *Styela plicata* (sea squirt); CHORDATA, MUSCLE-TYPE ISOFORMS (m): 47. act2_xentr *Xenopus tropicalis* (western clawed frog), 48. actc_human *H. sapiens,* 49. act1_xenla *Xenopus laevis* (African clawed frog), 50. act2_xenla *X. laevis,* 51. acts_human *H. sapiens,* 52. act3_xenla *X. laevis,* 53. acta_chick *Gallus gallus* (chicken), 54. acta_human *H. sapiens,* 55. acta_bovin *Bos taurus* (bovine), 56. acth_human *H. sapiens,* 57. actm_halro *Halocynthia roretzi* (sea squirt), 58. actm_stycl *Styela clava* (sea squirt), 59. actm_stypl *S. plicata;* NEMATODA: 9. act_caeel *Caenorhabditis elegans,* 10. act2_caeel *C. elegans,* 12. act1_oncvo *Onchocerca volvulus,* 13. act2_oncvo *O. volvulus;* ARTHROPODA: 11. act3_limpo *Limulus polyphemus* (Atlantic horseshoe crab) (I), 14. act4_artsx *Artemia* sp. (brine shrimp) (I), 15. act2_drome (c) *Drosophila melanogaster* (fruit fly) (I), 16. act1_drome (c) *D. melanogaster* (I) 17. act3_bommo (c) *Bombyx mori* (silk moth) (I), 21. acta_limpo *L. polyphemus* (II), 22. acty_limpo *L. polyphemus* (II), 23. act1_artsx *Artemia* sp. (I), 24. act2_artsx, *Artemis* sp. (I), 25. act5_drome (m) *D. melanogaster* (I), 26. act1_bommo (m) *B. mori* (I), 27. act2_bommo (m) *B. mori* (I), 28. act6_drome (m) *D. melanogaster* (I), 29. act4_drome (m) *D. melanogaster* (I); ECHINODERMATA: 18. actc_pisoc (c) *Pisaster ochraceus* (sea star), 19. actm_pisoc (m) *P. ochraceus,* 35. act2_strfn (c) *Strongylocentrotus franciscanus* (sea urchin), 36. actc_strpu (c) *Strongylocentrotus purpuratus* (purple sea urchin), 37. act1_strfn (c) *S. franciscanus,* 38. actm_strpu (m) *S. purpuratus,* 39. act3_strpu (c) *S. purpuratus;* MOLLUSCA: 20. act_aplca *Aplysia californica* (California sea hare); CNIDARIA (HYDROZOA): 32. act1_podca *Podocoryne carnea,* 33. act3_podca *P. carnea,* 34. act_hydat *Hydra attenuata* (hydra); PROTISTA: 69. act_enthi *Entamoeba histolytica,* 92. act1_naefo *Naegleria fowleri* (sarcomastigophora); PROTISTA-I: 84. act1_plafa *Plasmodium falciparum* (sporozoa), 85. act_crypv *Cryptosporidium parvum* (sporozoa), 93. act_tetpy *Tetrahymena pyriformis* (ciliate), 94. act_tetth *Tetrahymena thermophila* (ciliate), 95. act2_plafa *P. falciparum,* 98. act1_trybb *Trypanosoma brucei brucei* (sarcomastigophora), 99. act2_trybb *T. brucei brucei,* 100. act_oxyno *Oxytricha nova* (ciliate), 101. act_ocxyfa *Oxytricha fallax* (ciliate), 102. act_eupcr *Euplotes crassus* (ciliate), (100–102 are referred to as Protista-I′); PROTISTA-II: 40. act1_dicdi *Dictyostelium discoideum* (slime mold), 41. act8_dicdi *D. discoideum,* 42. act_phypo *Physarum polycephalum* (slime mold), 43. act2_dicdi *D. discoideum,* 44. act3_dicdi *D. discoideum,* 45. act1_acaca *Acanthamoeba castellanii* (amoeba), 90. act4_dicdi *D. discoideum,* 97. actd_phypo *P. polycephalum;* FUNGUS-I: 86. act1_phyin *Phytophthora infestans* (potato late blight fungus), 87. act_achbi *Achlya bisexualis,* 88. act2_phyin, *P. infestans,* 89. act_phyme *Phytophthora megasperma* (potato pink rot fungus); FUNGUS-II: 63. actg_emeni *Emericella nidulans,* 64. act-_thela *Thermomyces lanuginosa,* 65. act_schpo *Schizosaccharomyces pombe* (fission yeast), 66. act2_absgl *Absidia glauca* (pin mold), 67. act_yeast *Saccharomyces cerevisiae* (baker's yeast), 68. act_canal *Candida albicans* (yeast); PLANTA: 70. act1_pea *Pisum sativum* (garden pea), 71. act2_pea *P. sativum,* 72. act7_soltu *Solanum tuberosum* (potato), 73. act1_soltu *S. tuberosum,* 74. act_tobac *Nicotiana tabacum* (common tobacco), 75. act5_soltu *S. tuberosum,* 76. act1_arath *Arabidopsis thaliana* (mouse-ear cress), 77. act1_orysa *Oryza sativa* (rice), 78. act2_soltu *S. tuberosum,* 79. act_volca *Volvox carteri* (green algae), 80. act2_orysa *O. sativa,* 81. act1_dauca *Daucus carota* (carrot), 82. act1_maize *Zea mays* (maize), 83. act3_orysa *O. sativa,* 91. act7_orysa *O. sativa,* 96. act2_dauca *D. carota.*

Results

Screening the D. dendriticum *cDNA Library for Actin Transcripts*

After screening the *D. dendriticum* cDNA library for actin sequences, 40 phage clones that hybridized strongly were selected for further investigation. Restriction mapping and partial sequencing of the isolated clones indicated that all of the cDNA inserts contained actin-encoding sequences representing six different genes. Plasmids, *pDidact1–pDidact6,* which contained the longest inserts of the different cDNAs, were totally sequenced (Fig. 1). One of them, *pDidact1,* has been characterized previously (Wahlberg et al. 1994). The corresponding genes were given the names *Didact1–6.*

Sequence Analysis of Protein Coding Regions

pDidact2 (1,257 bp), *pDidact3* (1,356 bp), and *pDidact4* (1,329 bp) contain a complete coding region from the initiation codon ATG to the stop codon TAA as well as 5′ and 3′ untranslated sequences. *pDidact5* (1,175 bp) and *pDidact6* (1,239 bp), on the other hand, were found to lack 15–18 and 7–10 nucleotides, respectively, at the 5′-end of the open reading frame, judging from comparisons to the other *D. dendriticum* actin cDNAs (Fig. 1).

The deduced amino acid sequences of the full-length open reading frames are 376 (*Didact2* and *Didact4*) or 377 (*Didact3*) amino acids long (Fig. 2) and the calculated molecular weights are 41,772, 41,744, and 41,912 Da, respectively. According to these sequences and the partial amino acid sequences derived from *Didact5* and *Didact6,* we could conclude that the actins of *D. dendriticum* can be divided into three distinct groups.

The members of the first group, *Didact1, -2,* and *-4,*

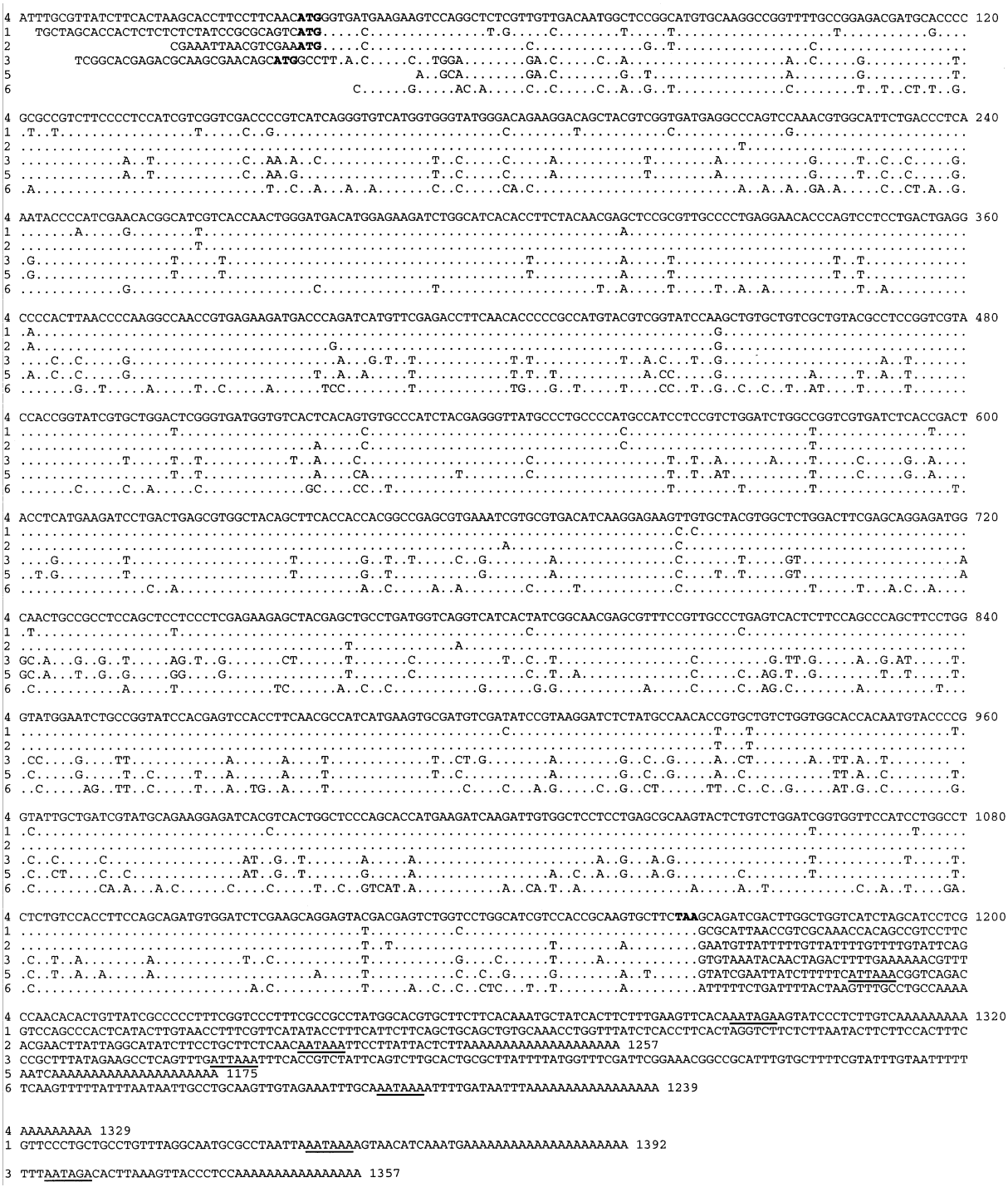

Fig. 1. Nucleotide sequences of *pDidact1* (*1*), *pDidact2* (*2*), *pDidact3* (*3*), *pDidact4* (*4*), *pDidact5* (5), and *pDidact6* (*6*) cDNAs from *D. dendriticum. Dots* indicate bases in the coding regions identical with *pDidact4.* The ATG start and TAA stop codons are typed in *bold* and putative polyadenylation signals are *underlined.* The *Didact2–Didact6* nucleotide sequences are available in the EMBL, GenBank, and DDJB databases under the accession numbers U27833–U27837.

are very similar. The putative *Didact1* protein is 100% identical with *Didact4*, and the deduced *Didact2* protein differs by only one amino acid from the other members of the group. On the nucleic acid level the identities range from 96.1 to 97.9% (Table 1). For the ease of data presentation these actins (*Didact1, -2,* and *-4*) are called the cestoda-I actin group. *Didact3* and *-5* encoded actins (cestoda-II actins) have 11 differing amino acids while the deduced peptide sequence of *Didact6* (cestoda-III actin) differs by more than 30 amino acids from any of the other *D. dendriticum* actin proteins (Table 1).

Fig. 2. Deduced amino acid sequences of the *pDidact1* and *pDidact4* (*4*), *pDidact2* (*2*), *pDidact3* (*3*), *pDidact5* (*5*), and *pDidact6* (*6*) cDNAs in *D. dendriticum.* Amino acids identical with *pDidact1/pDidact4* are indicated by *dots.*

Table 1. Pairwise comparisons of actin sequences from the cestodes *D. dendriticum, T. solium,* and *E. granulosus*[a]

	D. dendriticum						*T. solium*		*E. granulosus*	
	Diadact1	*Didact2*	*Didact3*	*Didact4*	*Didact5*	*Didact6*	AT5	AT6	*Egact1*	*Egct2*
Didact1		96.1	84.2	96.1	85.1	82.4	89.5	89.8	79.4	77.5
Didact2	99.7		84.8	97.9	85.3	82.7	88.9	89.7	79.9	77.9
Didact3	92.8	92.6		85.0	93.5	79.3	82.8	82.8	78.5	78.3
Didact4	100	99.7	92.8		85.5	82.7	89.2	89.6	79.8	78.1
Didact5	95.4	95.1	97.0	95.4		79.7	82.8	82.5	78.0	78.3
Didact6	91.4	91.4	90.9	91.4	90.8		80.4	81.0	80.7	76.5
AT5	100	99.7	92.8	100	95.4	91.4		97.7	79.4	77.9
AT6	100	99.7	92.8	100	95.4	91.4	100		80.1	77.9
Egact1	88.0	87.7	88.5	88.0	89.2	93.6	88.0	88.0		79.1
Egact2	88.3	88.0	89.6	88.3	90.6	86.3	88.3	88.3	85.6	

[a] The identity percentages between the coding region of the nucleic acid sequences is shown above the diagonal and the amino acid identity percentages below the diagonal. The AT5 and AT6 genes in *T. solium* code for identical proteins (act_taeso), while *Egact1* and *-2* correspond to the actin proteins act1_echgr and act2_echgr, respectively.

5′ and 3′ Untranslated Regions

The 3′ untranslated regions (UTRs) of *pDidact1–6* (Fig. 1) are 38–207 nucleotides long and they share 36.7–50.5% identity. In *pDidact2* and *-6* the conventional polyadenylation signal, AATAAA, is found 15 and 13 nucleotides, respectively, upstream of the poly(A) tail. The potential signal for poly(A) addition in *pDidact5* is ATTAAA, located 13 bp from the polyadenylation site, whereas only an unusual variant AATAGA is found close (18 and 15 nucleotides) to the poly(A) site of *pDidact3* and *-4*, respectively. In the 3′ UTR of *pDidact3* on ATTAAA hexanucleotide is located 117 nucleotides from the poly(A) tail. It is, however, not clear whether this sequence can function as an additional polyadenylation signal of the *Didact3* transcript. The 5′ UTRs of *pDidact1–4*, the cDNAs containing a 5′ UTR, are 16–35 nucleotides long (Fig. 1) and share 37.5–48.0% identity.

Phylogenetic Studies of Actin Amino Acid Sequences

In an attempt to dissect the relative relationships among the actins and thereby resolve the position of the *D. dendriticum* sequences, we have considered the use of principal components analysis (PCA) to help define clusters, followed by the averaging of members of each cluster to provide a single average representative to all other clusters. The net effect is to reduce the original 102 sequences to 14 clusters of sequences, including several individuals. In Fig. 3 are shown the two best 2D projections obtained by applying PCA to the 102 pairwise distances, *D*, derived from the multiple sequence alignment.

The actin clusters and individuals are as follows (Fig. 3, a complete listing is found in the Materials and Methods section): a group of plant actins, except for the *Volvox* actin, which clusters by itself, two major sets of protist actins with sequences from *Dictyostelium*, *Acanthamoeba*, and *Physarum* (act_phypo) in one cluster (protista-II actins); and two overlapping actin sets including sequences from ciliates, sporozoans, and *Trypanosoma* making up the other (protista-I). Actins from the protists *Entamoeba* (act_enthi), *Physarum* (actd_phypo), and *Naegleria* (act1_naefo) grouped by themselves as individuals. Two sets of fungal actins were seen: One set includes *Phytophthora* and *Achlya* se-

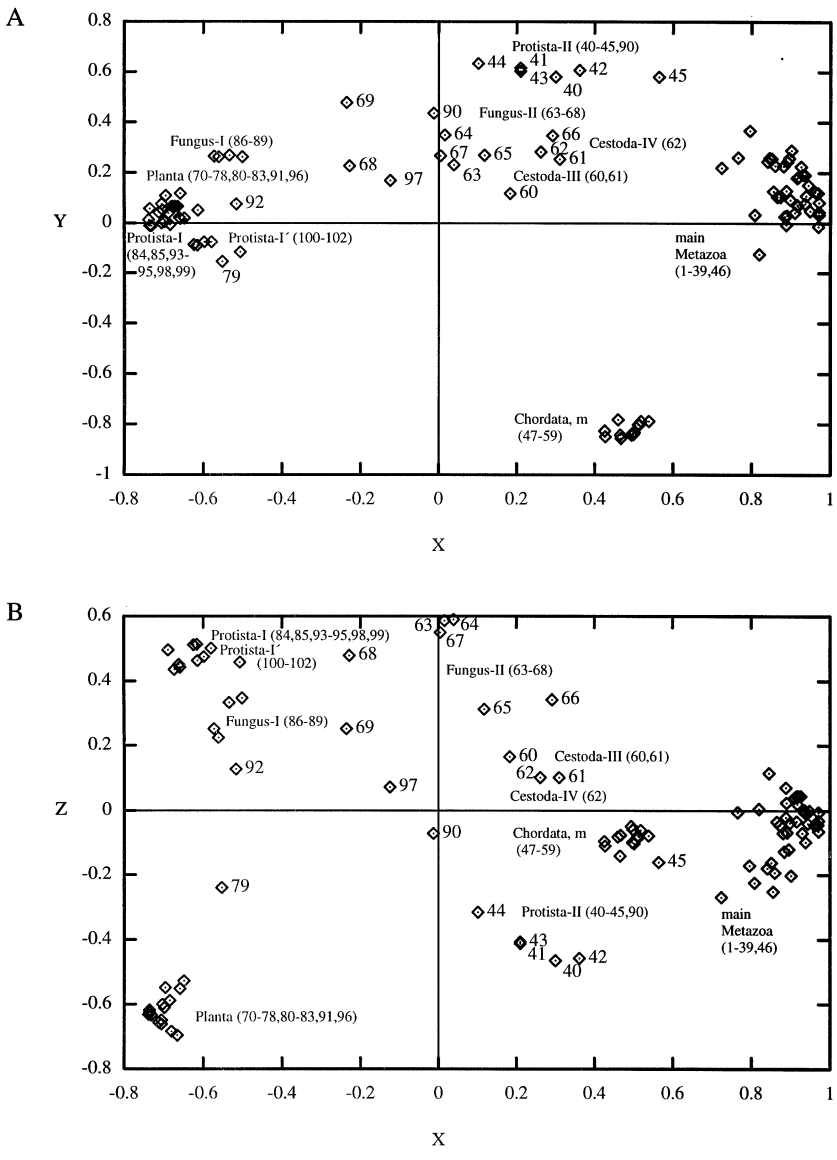

Fig. 3. Principal components analysis of pairwise distances obtained from the amino acid sequence alignment of 102 actin proteins. In **A** the projection of the two most significant dimensions (x = 48.4% of the total variance; y = 13.3%) and in **B** the next best 2D projection (x = 48.4%; z = 11.2%) show the clusters found to be within a Euclidean distance of 0.65 of each other when 11 dimensions are considered. The number codes and groups segregated in this manner are detailed in the Materials and Methods section. The actins not clustering with other sequences are numbers *69* (act_enthi), *79* (act_volca), *92* (act1_naefo), and *97* (actd_phypo).

quences (the fungus-I actin group) and the other, the fungus-II group, comprises the rest of the fungi sequences, i.e., yeast actins. The chordate muscle actins cluster together as do the major part of the remaining metazoan actins. This large group contains, among others, the actin proteins corresponding to the cestode sequences *Didact1–5* as well as the *Taenia* sequences. The *Didact6* amino acid sequence groups together with act1_echgr from *Echinococcus* (cestoda-III actin group) while the act2_echgr actin (cestoda-IV) clusters as an individual separated from other sequences (Fig. 3, see also Table 1 for cestode actin comparisons).

Distances between these clusters and individuals (averages of *D*) were then employed in the reconstruction of the phylogenetic relationships among the actins (Fig. 4). The two consensus trees made with the computer programs FITCH and NEIGHBOR show similar results except for the position of the *Entamoeba* actin, act_enthi. The NEIGHBOR tree is shown in Fig. 4. According to these results the two cestode acting groups, cestoda-III and -IV, branch off on the border between the other metazoan sequences and the fungus-protista-plant actin groups. To define the relationships of the four cestode sequences within the main metazoan actin group to the other sequences, it was necessary to consider this group in more detail.

Based on the clusters defined from the PCA clustering, the 40 sequences within the main metazoan actin group were clustered into 11 subgroups (not shown). These include two sets of cestode actins with the *Didact1* (=*Didact4* and act_taeso) and *Didact2* amino acid sequences in the first (cestoda-I actins) and the *Didact3* and *Didact5* sequences (cestoda-II) in the second set. Other clusters include the nematode actins, the arthropod actin proteins except acta_limpo and acty_limpo from the horseshoe crab *L. polyphemus*, which made up a group of their own, the vertebrate cytoplasmic actins, and Cnidaria sequences. The echinoderm actins were

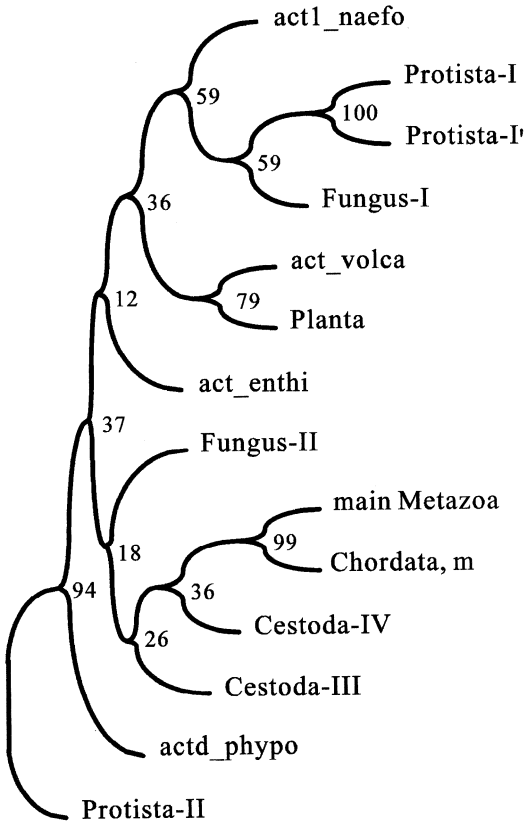

Fig. 4. Consensus unrooted tree constructed from group-average sequence-based distances for clusters defined by PCA (Fig. 3) using the program NEIGHBOR (Felsenstein 1985) and 1,000 bootstrap data sets. The *numbers* indicate the percentage of replicate trees which contained these branchpoints. See Materials and Methods for identification of the group members.

grouped into two sets representing the muscle-type and the cytoplasmic-type actins. Individuals not clustered with any other group are the *Aplysia* actin and the sea squirt cytoplasmic sequence actc_stypl.

Combining the 13 actin groups of Fig. 3 with the 11 metazoan actin subgroups thus led to 24 individual sequences and clusters for which sequence-based distances were averaged to all outliers (Table 2). The two resultant trees produced from this approach with the computer programs NEIGHBOR and FITCH are shown in Fig. 5A and B. The four groups of cestode actins map to two different regions of the trees. The cestoda-I set is closest to the *Aplysia* actin within the large main metazoan actin cluster, while the other three cestode actin groups lie on the border between the fungus-protista-plant clusters and the metazoan sequences. The group average percentage identities and sequence-based distances are found in Table 2.

For comparison a phylogenetic tree based on the 102 individual amino acid sequences was made (tree not shown). The same major branches were obtained and also in this tree the cestode actins were seen to map to two different regions, one on the border to the metazoan actins and the other within this group.

Discussion

The data presented here indicates that cDNAs representing six different actin genes have been isolated from a *D. dendriticum* cDNA library. The putative amino acid sequences are 90.8–100% identical and, together with other cestode actins isolated today, they can be divided into four groups. The actins of these different groups are extremely divergent in comparison with actins within other metazoan phyla.

Untranslated Regions

There is virtually no detectable sequence identity between the 5′ and 3′ UTRs of *D. dendriticum* cDNAs, respectively (Fig. 1). This is the case also when comparing *D. dendriticum* UTRs, with 5′ and 3′ noncoding regions of actin genes from other species (data not shown). Usually the sequence conservation in the untranslated regions of actin transcripts is very low (He and Haymer 1994), but, for example, the 3′ UTRs of the *Strongylocentrotus purpuratus* CyIIIa and CyIIIb actins share 78% identity (Flytzanis et al. 1989), and five *Halocynthia roretzi* actin genes are 82–100% identical in their 3′ noncoding region (Kusakabe et al. 1992). High identities between the UTRs of different genes may reflect a recent gene duplication (Kusakabe et al. 1992).

The Molecular Evolution of Actin

Defining the mutual relationships among the actin protein or nucleic acid sequences is a notorious problem in phylogenetic reconstructions of actin evolution (Mounier et al. 1992; Hennessey et al. 1993; Sheterline et al. 1995). Primarily, this is a result of the high sequence similarity among the actins (see Table 2); thus, relationships can be confused when a sequence is slightly more conserved or acquires several additional changes relative to the others. The end result is that a number of equally likely trees are obtained and the significance of nearly all major branchpoints is low (Reece et al. 1992; Hennessey et al. 1993; Sheterline et al. 1995). Because of uncertainties that occur within phylogenetic trees, we have chosen to score amino acid replacements according to a matrix customized for this collection of highly similar sequences and to cluster together sequences that clearly form closely related groups.

Based on the 2D projections from the sequence-distance matrix, the actin proteins that cluster near each other (Fig. 3) are in good agreement with the more distinct branches of phylogenetic trees published previously (Hennessey et al. 1993; Sheterline et al. 1995). The protist actins often cluster into two main groups (Hennessey et al. 1993; Sheterline et al. 1995). The definitions of the two protist actin groups in this study do not correlate with the definitions made by Reece et al. (1992) where the protista-I actins are labeled Pro2, and vice versa. Naming the groups is simply for the ease of data presentation and since the ciliate-sporozoa-*Trypanosoma* actins are more distant, in terms of their sequences, from all of

Table 2. Group-average percentage identities (right triangle); group-average sequence-based distances (averages of *D*) (left triangle); see Materials and Methods for identification of group members

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.
1. Cestoda-I		95.9	95.5	95.6	94.1	94.6	94.1	92.7	93.3	92.9	93.8	92.3
2. Vertebrata, c	2.7		96.8	96.1	96.2	96.1	95.2	94.0	95.4	94.6	95.0	93.4
3. Nematoda	2.7	2.1		96.3	95.5	95.9	94.8	94.5	95.4	94.4	94.8	93.8
4. Aplysia	2.9	2.7	2.2		94.8	95.3	94.0	94.0	95.3	93.6	95.7	92.8
5. Arthropoda-II	3.6	2.4	2.5	3.3		95.1	94.2	93.3	95.5	94.0	94.6	91.9
6. Arthropoda-I	3.3	2.6	2.4	3.0	2.9		94.0	93.6	94.8	93.5	94.5	92.8
7. Cestoda-II	3.9	3.4	3.5	4.1	3.9	4.1		92.5	94.4	92.6	94.1	91.5
8. Echinodermata, c	4.7	4.0	3.5	3.9	4.2	4.2	5.2		94.5	92.3	94.3	92.3
9. Echinodermata, m	4.2	3.1	2.8	3.1	2.9	3.3	4.1	3.6		94.0	96.2	92.3
10. Styela, c (Tunicata)	4.9	3.7	4.0	4.7	4.3	4.7	5.4	5.6	4.4		94.5	90.8
11. Cnidaria	3.9	3.3	3.2	2.9	3.4	3.6	4.2	3.8	2.5	4.1		92.3
12. Chordata, m	5.2	4.9	4.3	5.0	5.5	5.1	6.1	5.6	5.4	6.8	5.4	
13. Cestoda-III	6.7	6.2	5.8	6.7	6.4	6.5	6.7	8.0	6.9	7.9	7.1	7.9
14. Cestoda-IV	7.5	6.8	6.9	6.7	6.6	7.3	6.9	8.1	7.1	8.4	6.8	9.3
15. Fungus-II	8.0	7.2	7.4	7.7	8.1	8.1	8.0	8.8	8.0	9.1	8.0	9.6
16. Protista-II	5.8	5.1	5.2	5.8	5.5	5.8	6.1	6.9	5.4	6.7	5.3	7.7
17. actd_phypo	16.0	15.3	15.3	15.1	16.0	15.7	15.8	16.7	15.4	17.0	14.9	16.9
18. act_enthi	7.3	7.5	7.5	8.5	7.9	7.9	8.6	8.9	8.0	8.4	8.0	10.3
19. Planta	9.7	9.5	9.5	9.6	9.5	9.9	10.0	10.4	9.4	11.5	9.1	10.3
20. act_volca	6.6	6.1	6.3	6.2	6.6	7.0	6.8	7.9	7.0	8.7	7.0	7.0
21. act1_naefo	13.7	14.4	13.5	13.8	13.8	14.1	14.7	14.3	14.2	16.3	13.9	15.1
22. Fungus-I	10.6	10.3	10.7	10.8	10.6	10.8	11.3	11.7	10.7	11.3	9.8	12.3
23. Protista-I	16.5	16.4	16.2	16.6	16.6	16.8	17.0	17.4	17.1	17.7	16.7	17.4
24. Protista-I′	29.3	28.5	28.9	28.2	28.9	29.0	29.1	29.2	28.9	29.8	28.7	28.8

Table 2. Continued

	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.
1. Cestoda-I	89.7	88.2	87.9	91.1	79.4	88.7	86.4	90.3	80.7	84.2	76.4	63.0
2. Vertebrata, c	90.6	89.7	89.0	92.8	80.4	88.8	87.1	91.2	80.0	84.6	76.9	64.1
3. Nematoda	91.4	89.0	88.6	92.2	80.2	88.8	86.5	90.3	81.0	83.8	76.8	63.4
4. Aplysia	90.0	89.6	88.4	91.6	80.9	87.5	86.8	91.0	80.8	84.0	76.4	64.3
5. Arthropoda-II	90.5	89.8	87.7	92.1	79.6	88.2	87.0	90.3	80.8	84.4	76.5	63.4
6. Arthropoda-I	90.3	88.5	87.8	91.4	80.1	88.1	86.3	89.6	80.2	84.0	76.0	63.6
7. Cestoda-II	90.0	90.1	88.3	91.5	79.4	87.6	86.5	90.2	79.6	83.7	76.5	63.4
8. Echinodermata, c	88.3	87.6	86.8	90.0	78.7	87.1	85.8	88.4	80.4	82.7	75.7	63.2
9. Echinodermata, m	89.9	89.4	87.8	92.4	80.5	88.1	87.0	89.6	80.4	84.2	75.8	63.6
10. Styela, c (Tunicata)	88.4	87.7	86.4	90.7	78.5	88.3	84.6	87.7	77.9	83.3	75.5	62.6
11. Cnidaria	89.5	89.5	87.6	92.0	80.9	87.9	87.2	89.6	80.8	85.0	76.2	63.8
12. Chordata, m	88.6	86.2	86.1	89.3	78.8	85.4	85.5	89.2	79.4	82.1	75.7	63.8
13. Cestoda-III		85.9	84.8	88.5	78.2	86.8	84.1	87.1	80.1	81.6	75.8	63.0
14. Cestoda-IV	9.5		83.5	86.5	76.3	83.0	83.1	85.4	77.3	80.4	74.3	61.9
15. Fungus-II	10.1	11.1		86.2	76.7	83.4	82.1	86.0	75.9	80.3	75.2	62.7
16. Protista-II	7.9	9.3	9.3		80.3	87.7	86.2	88.5	79.5	83.6	75.2	63.2
17. actd_phypo	16.9	18.6	17.9	15.4		76.8	75.7	77.1	72.2	72.2	67.4	57.9
18. act_enthi	9.2	11.7	11.3	8.3	17.9		83.2	84.8	78.7	81.4	76.3	64.5
19. Planta	11.3	12.1	12.7	9.8	19.2	12.0		87.2	78.3	81.1	74.8	62.3
20. act_volca	8.7	9.7	9.5	7.9	17.8	10.6	8.7		80.0	84.2	77.4	63.7
21. act1_naefo	14.4	16.3	17.3	14.8	22.3	15.6	16.2	14.3		77.5	72.4	59.6
22. Fungus-I	12.5	13.3	13.0	11.0	21.4	12.6	13.3	10.7	16.5		75.7	62.6
23. Protista-I	17.5	18.1	17.8	17.6	25.9	17.2	18.5	15.9	20.8	17.3		62.5
24. Protista-I′	29.2	30.3	29.8	29.3	35.0	28.1	30.6	28.8	33.1	29.7	30.1	

the other actins than are the other protist actins in the phylogenetic studies, we labeled these actins protista-I sequences. Note, however, that the trees presented here are unrooted.

The most distant of the metazoan actins, assuming that the root of the tree depicted in Fig. 4 lies along a branch within the nonmetazoan actins (i.e., in the vicinity of the protista-I/fungus-I/plant groups), are the cestode actins group III (*Didact6*, act1_echgr) and IV (act2_echgr). The chordate muscle actins, represented by

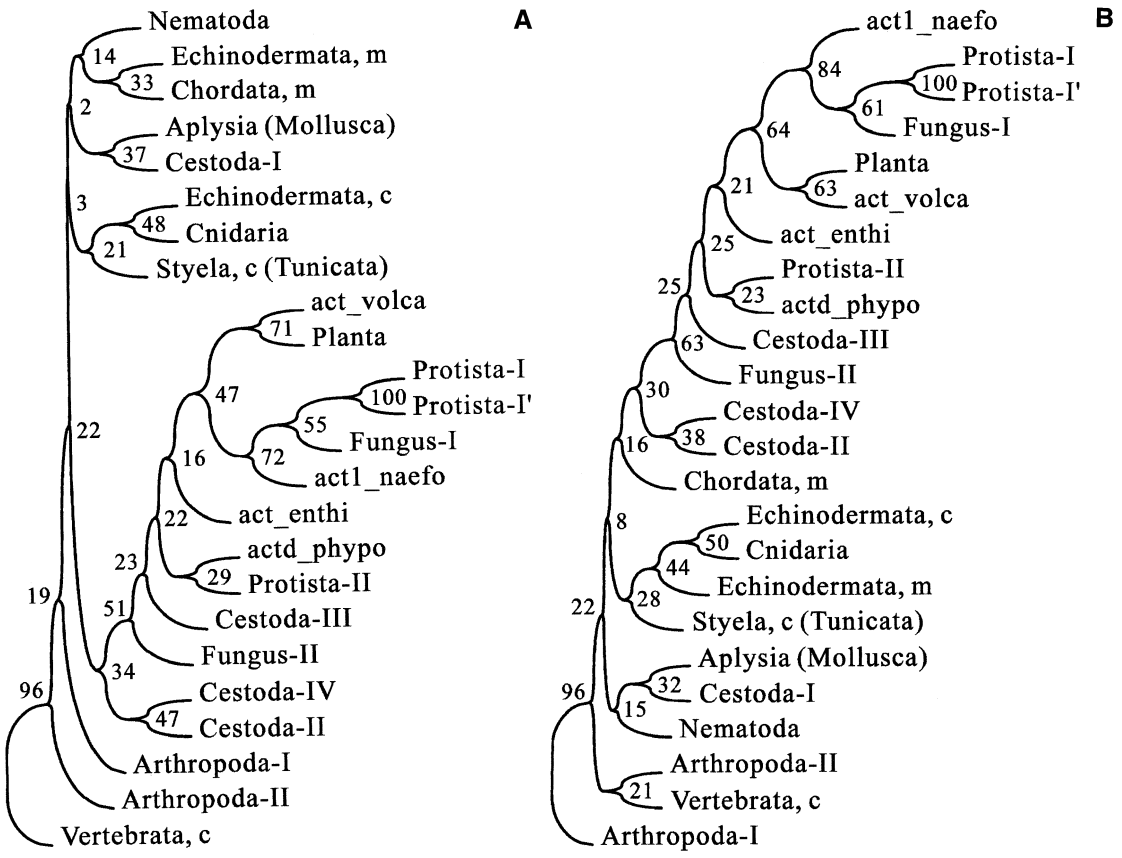

Fig. 5. Consensus unrooted trees constructed from group-average sequence-based distances for clusters defined by PCA (Fig. 3) plus additional clusters defined within the main metazoan actin group. The *numbers* are percentages of 1,000 bootstrap replicates where the same internal branch is recovered. Trees were constructed using the programs **(A)** NEIGHBOR and **(B)** FITCH (Felsenstein 1985). See Materials and Methods for identification of the group members.

actins from vertebrates and tunicates (urochordates), are also well separated into a group apart from all the remaining metazoan actins (see also Fig. 3).

When the main metazoan actin group was studied in more detail it was found to include two additional sets of cestode actins (Fig. 5A,B). One group, cestoda-I, containing *Didact1, -2,* and *-4* plus the act_taeso proteins, clusters together with the *Aplysia* actin. This is the case also in other studies where the cestode actin (act_taeso) is included (Sheterline et al. 1995; Hennessey et al. 1993; Mounier et al. 1992). The other group of cestode actins in the main metazoan actin group, cestoda-II containing *Didact3* and *-5,* appears to branch off close to the cestoda-III and -IV actin groups early in metazoan evolution.

In the work of Mounier et al. (1992), Hennessey et al. (1993), and Sheterline et al. (1995) the cnidaria actins cluster together with, or close to, the echinoderm nonmuscle actins (see also the NEIGHBOR tree in Fig. 5A of this study). The echinoderm muscle specific isoforms, on the other hand, are seen to group together with chordate muscle actins, and this cluster is branching off after the echinoderm-cytoplasmic-cnidaria actin group. These results are in agreement with the theory of Kovilur et al. (1993) according to which the evolution of the chordate muscle actin genes, from a non–muscle-actin sequence, probably begun early in the protochordate lineage, maybe in an early echinoderm-like ancestor. The emergence was associated with an increased rate of molecular evolution (Mounier et al. 1992). According to the phylogenetic FITCH tree of the present study (Fig. 5B), however, the cnidaria actins cluster with both the echinoderm cytoplasmic and muscle actin groups, the chordate muscle actins branching off before this cluster. This controversy could be due to methodological problems because of an increase in the rate of molecular evolution within the chordate muscle actin lineage in comparison with the high sequence conservation among actins overall. Indeed, when a small change in rate is applied on the chordate muscle actins in Fig. 5B, they group together with echinoderm muscle actin sequences (data not shown).

As can be seen from the bootstrap values in this and other studies (Hennessey et al. 1993; Mounier et al. 1992; Sheterline et al. 1995), the major branches of the actin phylogenetic trees cannot be considered significant. This is due to the overall high sequence similarity among the actin proteins coupled with variable rates of change among different groups. This makes it very difficult to draw absolute conclusions about the phylogenetic posi-

tion of the cestode actins. What we do know with certainty is that the cestode actins can be divided into at least four groups, possibly reflecting different cellular and/or tissue functions (Fig. 5A,B). At the moment, *D. dendriticum* has members in three of the groups and *E. granulosus* in two. So far, genes coding for only one *T. solium* actin protein have been isolated, and the question remains as to whether the cestode species have representatives in each of the four groups.

We also know that these different cestode actin types are extremely divergent from each other. The only other group among the metazoan actins where the members differ to such an extent is the vertebrate actin group containing both muscle and cytoplasmic isoforms. Since the cestode actins differ from each other to such a degree, and because no cestoda-III or -IV actin types have been found in other organisms so far, it is probable that the changes have appeared within the cestode actin group as the result of an increasing rate of molecular evolution compared to actins overall. Whether the progenitor cestode actin gene is an ancestral cestoda-I actin-like sequence or whether the cestode actins phylogenetically branched off early in metazoan evolution is, according to the available data, difficult to say.

According to the hypothesis of Barnes et al. (1988), ancestral flatworm-like animals are considered to be ancestors for the metazoans living today with the exception of the cnidarians, ctenophora, porifera, and placozoa. Also in other studies, based on structural or sequence (18S rRNA) comparisons, the cnidarians and platyhelminths are among the first groups to branch off during the metazoan evolution (Adoutte and Philippe 1993; Field et al. 1988).

In conclusion, the results show that *D. dendriticum* contains at least six different genes coding for actin in its genome. According to phylogenetic studies it is clear that the cestode actins known today cluster together forming four distinct groups, in three of which *D. dendriticum* has members at the moment. The different types of cestode actins are very different from each other in comparison with metazoan actins in general. These differences might be due to changes of the evolutionary rate within the cestode actin group, which makes the absolute phylogenetic positioning of these actins difficult.

Acknowledgments. This work was supported by Victoriastiftelsen, Åbo Akademi University, and the Academy of Finland. The authors thank M.Sci. Kaj Karlstedt for valuable ideas and discussions and Prof. Pertti Panula for reading the manuscript.

References

Adoutte A, Philippe H (1993) The major lines of metazoan evolution: summary of traditional evidence and lessons from ribosomal RNA sequence analysis. In: Pichon Y (ed) Comparative molecular neurobiology. Birkhäuser Verlag, Basel, Switzerland, pp 1–30

Barnes RSK, Calow P, Olive PJW (1988) The invertebrates: a new synthesis. Blackwell Scientific, Oxford

Benton WD, Davis RW (1977) Screening γ-gt recombinant clones by hybridization to single plaques *in situ.* Science 196:180–182

Chatfield C, Collins AJ (1989) Introduction to multivariate analysis. Chapman and Hall, New York

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Drouin G, de Sá MM, Zuker M (1995) The *Giardia lamblia* actin gene and the phylogeny of eukaryotes. J Mol Evol 41:841–849

Fang H, Brandhorst BP (1994) Evolution of actin gene families of sea urchins. J Mol Evol 39:347–356

Felsenstein J (1985) Confidence limits on phylogenesis: an approach using the bootstrap. Evolution 39:783–791

Field KG, Olsen GJ, Lane DJ, Giovannoni SJ, Ghiselin MT, Raff EC, Pace NR, Raff RA (1988) Molecular phylogeny of the animal kingdom. Science 239:748–753

Flytzanis CN, Bogosian EA, Niemeyer CC (1989) Expression and structure of the CyIIIb actin gene of the sea urchin *Strongylocentrotus purpuratus.* Mol Reprod Dev 1:208–218

He M, Haymer DS (1994) The actin gene family in the oriental fruit fly *Bactrocera dorsalis.* Muscle specific actins. Insect Biochem Mol Biol 24:891–906

Hennessey ES, Drummond DR, Sparrow JC (1993) Molecular genetics of actin function. Biochem J 291:657–671

Hightower RC, Meagher RB (1986) The molecular evolution of actin. Genetics 114:315–332

Hori H, Osawa S (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. Mol Biol Evol 4:445–472

Johnson MS, Overington JP (1993) A structural basis for the comparison of sequences: an evaluation of scoring methodologies. J Mol Biol 233:716–738

Johnson MS, Overington JP, Blundell TL (1993) Alignment and searching for common folds using a data bank of structural templates. J Mol Biol 231:735–752

Kovilur S, Jacobson JW, Beach RL, Jeffery WR, Tomlinson CR (1993) Evolution of the chordate muscle actin gene. J Mol Evol 36:361–368

Kusakabe T, Makabe KW, Satoh N (1992) Tunicate muscle actin genes. Structure and organization as a gene cluster. J Mol Biol 227:955–960

Maeda N, Smithies O (1986) The evolution of multigene families: human haptoglobin genes. Annu Rev Genet 20:81–108

Meagher RB (1991) Divergence and differential expression of actin gene families in higher plants. Int Rev Cytol 125:139–163

Miwa T, Manabe Y, Kurokawa K, Kamada S, Kanda N, Bruns G, Ueyama H, Kakunaga T (1991) Structure, chromosome location, and expression of the human smooth muscle (enteric type) γ-actin gene: evolution of six human actin genes. Mol Cell Biol 11:3296–3306

Mounier N, Gouy M, Mouchiroud D, Prudhomme JC (1992) Insect muscle actins differ distinctly from invertebrate and vertebrate cytoplasmic actins. J Mol Evol 34:406–415

Reece KS, McElroy D, Wu R (1992) Function and evolution of actins. In: Hecht MK, Wallace B, MacIntyre RJ (eds) Evolutionary biology. Plenum Press, New York, 26:1–34

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Sheterline P, Clayton J, Sparrow JC (1995) Actin. Protein Profile 2(1): 1–103

Vandekerckhove J, Weber K (1978) Actin amino-acid sequences. Comparison of actins from calf thymus, bovine brain, and SV40-transformed mouse 3T3 cells with rabbit skeletal muscle actin. Eur J Biochem 90:451–462

Wahlberg MH, Karlstedt KA, Paatero GIL (1994) Cloning, sequencing and characterization of an actin cDNA in *Diphyllobothrium dendriticum* (Cestoda). Mol Biochem Parasitol 65:357–360, Addendum 68:334