

Debashish Bhattacharya · Klaus Weber

The actin gene of the glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes

Received: 4 December 1996 / Accepted: 5 February 1997

Abstract We isolated the actin gene of the glaucocystophyte alga *Cyanophora paradoxa* and analyzed the coding region and its introns. Phylogenetic analyses of the actin coding region and the inferred protein sequence in data sets containing 47 other actin sequences show *Cyanophora* to be a member of the eukaryotic crown-group radiation in agreement with ribosomal DNA sequence analyses. Four of the five *Cyanophora* actin introns are relatively short (55–59 nt) and occupy novel positions in a catalogue of actin introns containing 56 distinct sites. The fifth intron has a length of 171 nt and occurs also in actin genes from green algae and the crustacean *Artemia*.

Key words Actin · *Cyanophora paradoxa* · Introns · Phylogeny

Introduction

The glaucocystophytes represent a small (nine genera, Kies 1979; Kies and Kremer 1990), morphologically distinct, protist group that is distinguished by its photosynthetic organelles, the cyanelles. Characters such as cruciate flagellar roots with associated multilayered structures (MLS) have suggested a phylogenetic relationship between glaucocystophytes and the green algae (Melkonian 1983), whereas cyanelle characters have been used to position glaucocystophytes with the red algae (Cavalier-Smith 1982). Cyanelles resemble red-algal plastids in that they contain only chlorophyll-*a* and phycobilisomes and have unstacked thylakoids (Kies and Kremer 1990). Cyanelles were previously thought to represent recent, possibly independent, endosymbioses because these organelles main-

tain the peptidoglycan wall of the cyanobacterium as a bounding plastid membrane. Phylogenetic analyses of nuclear-encoded small subunit ribosomal DNAs (SSU rDNAs) identify the glaucocystophytes as a member of the eukaryotic “crown-group” radiation and a sister group to the cryptophyte algae (with weak bootstrap support, see Bhattacharya and Medlin 1995). Glaucocystophytes do not share a specific relationship with red or green algae in rDNA phylogenies (Bhattacharya et al. 1995). Analyses of the cyanelle-encoded rDNAs show these organelles to be “advanced” plastids that form a monophyletic group closely related to red- and green-algal plastids (Helmchen et al. 1995).

To gain another perspective on the phylogeny of the glaucocystophytes, we have determined the sequence of the single-copy actin gene of *Cyanophora paradoxa* and included it in phylogenies based on actin sequences from primitive eukaryotes as well as members of the crown groups. Actin is a highly conserved eukaryotic protein that can be used in phylogenetic analyses to corroborate results based on rDNA and other protein-sequence analyses (Bhattacharya et al. 1991, 1993; Baldauf and Palmer 1993; Bouget et al. 1995; Drouin et al. 1995). Previous actin trees have, for example, been consistent with a sister-group relationship between animals and fungi proposed on the basis of rDNA sequence analyses (Baldauf and Palmer 1993; Wainright et al. 1993), and an early divergence of the archazoan *Giardia l.* from the eukaryote phylogeny (Sogin et al. 1989; Baldauf and Palmer 1993; Drouin et al. 1995). Actin trees have also sometimes conflicted with the results of rDNA sequence comparisons (e.g. the position of slime molds, see Bhattacharya et al. 1991; Drouin et al. 1995 and below) and provided the impetus to probe more deeply into the resolving power of rDNA and protein-sequence phylogenies (Loomis and Smith 1990). For these reasons, actin-sequence comparisons are an important addition to analyses of eukaryote/gene phylogeny, and in those cases where actin exists as a single-copy coding region (e.g. fungi, ciliates, most green algae) it is taken as a working hypothesis that these sequences may be used as a molecular marker. In other lineages where actin genes form multi-

D. Bhattacharya (✉) · K. Weber
Department of Biochemistry,
Max Planck Institute for Biophysical Chemistry,
Am Fassberg 11, D-37077 Göttingen, Germany

Communicated by R. Schweyen

copy families (e.g. land plants, metazoa) these sequences can be used to study aspects of concerted gene evolution and, within orthologous subfamily members, may also be used as phylogenetic markers (Moniz de Sá and Drouin 1996).

The analyses of spliceosomal intron positions is also well established with actin genes due to the highly conserved sequence and length of most of this protein (except for the variable N-terminus) that allow for the unambiguous alignment of intron sites. The existence of a putative proto-splice site bordering intron insertion positions [i.e. (^{C/A})AG-intron-(^{A/G})] was, for example, recognized on the basis of actin and tubulin gene analyses (Dibb and Newman 1989). In the present paper we have incorporated the *Cyanophora* actin introns into an updated catalogue of actin spliceosomal intron positions among all eukaryotes (see also Weber and Kabsch 1994).

Materials and methods

Isolation of the actin DNA clone. Genomic DNA was isolated from an actively growing culture of *C. paradoxa* Korsch. [Kies strain, Sammlung von Algenkulturen Göttingen (Schlösser 1984), SAG B 45.84] as previously described (Helmchen et al. 1995). A genomic DNA bank was established from this taxon using the *Bam*HI-digested, de-phosphorylated vector Lambda-Dash (Stratagene) and partially digested *Bam*HI genomic DNA from *Cyanophora*. Ligation and packaging of the vector and DNA were done according to the manufacturer's instructions (Stratagene). The bank was amplified (50 000 original PFU) and stored in SM buffer [10 mM NaCl, 10 mM MgCl₂, 50 mM Tris-HCl (pH 7.5), 2% w/v gelatin] with 5% chloroform. The genomic bank of *Cyanophora* was screened with an actin-encoding cDNA PCR fragment isolated from this taxon in 6×SSC, 30% formamide and 5×Denhardt's solution at 40°C, overnight. The actin probe was isolated with a RT/PCR method using the forward primer 244F (5'-GAYATGGAAAAGATCTGG-3') and a universally conserved reverse primer at the carboxy terminus of actin-coding regions (DBAc3, 5'-CCGAATTCTGCAGARAAG-CAYTTS CTGTG-3'; see Bhattacharya et al. 1991, 1993). Filters were washed twice in 2×SSC, 0.1% SDS at 40°C. Actin-containing phages were isolated and a *Xba*I fragment (of size 2.2 kb) from one of these phages, that cross-reacted with the *Cyanophora* RT/PCR actin fragment in a Southern analysis, was subcloned into the *Xba*I restriction site in pBluescript for sequencing. Actin sequences were determined over both strands of this gene using primers complementary to conserved regions within all actin-coding regions (Bhattacharya et al. 1991), or by primer-walking using *Cyanophora*-specific actin oligonucleotides and the dideoxy sequencing protocols (Sanger et al. 1977) with a T7 sequencing kit (Pharmacia). The *Xba*I fragment characterized encodes the actin genomic sequence lacking approximately 57 nt at the 3' end of the coding region. The remaining genomic actin sequence of *Cyanophora* could not be found within the original Lambda-Dash clone. To extend the 3'-terminus of the actin-coding region, the RT/PCR method was used with *Cyanophora* total RNA extracted with an RNeasy kit (Qiagen). First-strand synthesis was initiated with an oligo-dT₁₂₋₁₈ primer (Pharmacia). The actin-coding region was amplified from this cDNA using the forward primer 613F (TTCACCACCACTGCTGAG-3') and the reverse primer, DBAc3. This PCR fragment was cloned into the *Sma*I restriction site in pBluescript and sequenced using the T3 and T7 primers. The actin gene sequence of *Cyanophora* is available from the EMBL/GenBank/DBJ databases under the accession number, U90325.

Southern analysis. Genomic DNA was prepared from *Cyanophora* as described above and digested to completion with the restriction

enzymes *Sal*I, *Sac*I and *Sma*I. The restriction products were size-fractionated in a 1% agarose gel in 1×TBE (89 mM Tris·Cl, 89 mM boric acid, 2 mM EDTA) electrophoresis buffer and transferred onto a nitrocellulose membrane (Schleicher and Schuell) and baked at 80°C for 2 h. The Southern blot was pre-hybridized at 40°C in a 6×SSC, 50% formamide, 5×Denhardt's and 0.1% SDS solution for 2 h and then incubated overnight in 5×SSC, 50% formamide, 1×Denhardt's and 0.1% SDS with the 613F-DBAc3 *Cyanophora* actin RT/PCR fragment as probe. The actin fragment was radioactively-labelled with ³²P-dCTP using a random-primed oligolabelling kit (Pharmacia). The Southern blot was washed twice for 1 h each time in 2×SSC, 0.1% SDS at 50°C prior to autoradiography. A further more-stringent wash was done in 0.1×SSC, 0.1% SDS at 55°C for 1 h with the same membrane to test the specificity of the actin-annealing fragments resolved in the first autoradiogram.

Phylogenetic analyses. The actin-coding region from *Cyanophora* was aligned with homologous sequences from 47 other eukaryotes and two contractin sequences (actin-related proteins; see Clark and Meyer 1992; Lees-Miller et al. 1992; Sheterline et al. 1995) from man and dog. Two data sets were prepared from this alignment. The first data set included the contractin sequences as the outgroup for eukaryotes ranging from *Giardia* to members of the crown group. This data set was used to determine the order of divergence of the most primitive eukaryotes in the analyses (e.g. *Giardia*, *Trypanosoma b.*) and their phylogenetic position relative to the crown-group radiation. The gene duplication that resulted in the actin and contractin genes is presumed to have occurred in the common ancestor of all eukaryotes (Bhattacharya and Ehling 1995; Sheterline et al. 1995). The second data set included a larger number of actin sequences from crown-group eukaryotes than in the first data set, with the apicomplexan (*Plasmodium f.*, *Toxoplasma g.*) actin-coding regions as an outgroup. The apicomplexans diverge prior to the crown-group radiation in actin phylogenies (see also Baldauf and Palmer 1993; Drouin et al. 1995). The second data set was used to infer the phylogenetic relationships of the crown-group lineages. The first and second nucleotide positions of actin codons (748 nt total) in both data sets were submitted to a weighted [re-scaled consistency (RC) index over the interval 1–1000] maximum-parsimony (MP) analysis (PAUP V3.1.1, Swofford 1993) using a heuristic search procedure with a branch-swapping algorithm (TBR, tree bisection-reconnection). The stability of internal nodes within the MP actin phylogenies was studied with a bootstrap analysis [100 replications (first data set), 200 replications (second data set), Felsenstein 1985]. Both data sets were also analyzed with a distance method using a Kimura (1980) matrix as input for a neighbor-joining phylogenetic reconstruction (Saitou and Nei 1987) with jumbled taxon addition and the transition/transversion ratio = 2 (Phylip V.3.5c, Felsenstein 1993). Bootstrap analyses (200 replications, Felsenstein 1985) were done with the distance method. Further, the phylogeny of the actin protein sequences in the second data set (372 amino acids) was studied using a distance method (PROTDIST, Dayhoff distance matrix, Phylip V.3.5c) and a neighbor-joining phylogenetic reconstruction. One-hundred bootstrap re-samplings were used to test the stability of the internal nodes in the protein distance tree. All actin sequences used in these analyses are available from the EMBL/GenBank/DBJ databases.

Results and discussion

Southern analysis of the *Cyanophora* actin-coding region

Southern analysis using the actin RT/PCR probe shows one actin-annealing band of approximate size 1100 nt in the *Sal*I restriction digest, and two bands of approximate sizes 1050 nt and 3000 nt in the *Pst*I digest of *Cyanophora* genomic DNA (Fig. 1). Restriction analysis of the *Xba*I fragment encoding the genomic *Cyanophora* actin sequence

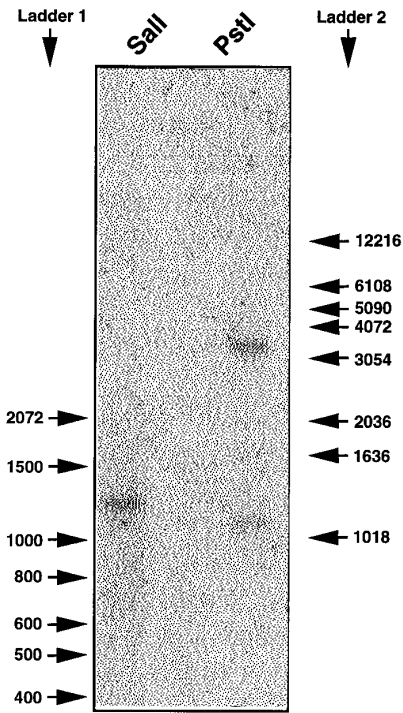


Fig. 1 Southern analysis of the *Cyanophora* genomic DNA with a homologous actin-cDNA probe. Ladders 1 (100 bp) and 2 (1 kb) are from BRL

shows an internal *PstI* restriction site 300 nt from the 3' terminus of this insert. A single band of approximate size 2.1 kb was resolved when *Cyanophora* genomic DNA digested with *SmaI* was included in the Southern analysis (data not shown). These results were also found after more stringent washing of the membrane (see Materials and methods). Sequence analysis (T3, T7 primers) of five independent clones isolated from the RT/PCR experiments (613F+DBAc3 PCR primers) with *Cyanophora* also showed identity indicating that there is one actin-encoding mRNA expressed in *Cyanophora*. The PCR primers are not likely to have bound to only one class of actin mRNAs since these oligonucleotides recognize highly conserved regions of the coding region. The results of the Southern and RT/PCR-sequence analyses are consistent with the existence of a single-copy actin gene in the genome of *Cyanophora*.

Evolutionary analysis of the *Cyanophora* actin gene

The actin gene of *Cyanophora* is interrupted by five spliceosomal introns that occur at positions 44-1, 248-3, 283-2, 318-3 and 345-2 relative to the vertebrate skeletal muscle α -actin sequence (Fig. 2). To determine if additional introns may be located in the 3'-terminal sequence that was determined from the cDNA, a PCR amplification

Fig. 2 The *Cyanophora* actin sequence determined from a genomic clone and the actin cDNA (see text for details). The coding region is shown in upper case whereas the five introns are shown in lower case. This sequence begins with the putative methionine start codon in actin and ends with the sequence of the PCR primer DBAc3 (shown in bold letters) used to isolate the cDNA. Nucleotides that could not be unambiguously determined are indicated with the letter N. The putative branch-point sequences within the introns are underlined and the consensus branch-point sequence with the branch point adenine in shadowed print is shown at the bottom of the figure

```

1 ATGGAAGACGAGGTTCCGGCGCTGGTCATTGACAAATGGCACCGGCATGTGCAAGGCCGGCTTCGCCGGCGATGCCGCCCGCTGCCGRTTCCCGGTCA
M--E--D--E--V--A--A--L--V--I--D--N--G--T--G--M--C--K--A--G--F--X--X--D--A--A--R--R--A--V--F--P--V--
101 TCGTGGGTGCCCGCGCCACCAAGgtccggtgtgtctgtagggtccgcattccgtgtgtccaaaactcgaaaaccgcctcaccgcctaataaaaccgtta
I--V--G--A--P--R--H--Q-----
201 cgctagacaagttatagctattagaggccttcaacgccggcgccttcggttcggttcgattttacttagcattcaaaaatggttgttctcggcagCGCTG
-----G--V-----
301 ATGGTCGCATGGGGCAGAAAGACTCGTATGTGCGCGACGAGCGCAGTCAAGCGTGGTATCCTCACGCTGAAGTACCCGATTGAGCAGCGCATCGTGA
-M--V--G--M--G--Q--K--D--S--Y--V--G--D--E--A--Q--S--K--R--G--I--L--T--L--K--Y--P--I--E--H--G--I--V--
401 CCAACTGGGACGACATGGAGAAGATTGGCACACACCTTCTACAACGAGCTCCCGGTTCGCCCGGAGGACCCCGTCTGCTCACTGAGGCGCGCT
T--N--W--D--D--M--E--K--I--W--H--H--T--F--Y--N--E--L--R--V--A--P--E--E--H--P--V--L--L--T--E--A--P--L
501 GAACCCGAAGGCCAACCCGCGAGAAGATGACCCAGATCATGTCGAGACCTTCAACGTGCCGGCGATCAGCTGCCATCCAGGCCGCTGCTCTCCCTCTAC
--N--P--K--A--N--R--E--K--M--T--Q--I--M--F--E--T--F--N--V--P--A--M--Y--V--A--I--Q--A--V--L--S--L--Y--
601 GCCTCCGCGCTACCCACCGTATCGTGTGCTCGACTCCGCGATGGTGTCAACCCACACCGTCCCATCTACGAGGGTTANGCGTCCCGCAGCCATCTCTCC
-A--S--G--R--T--T--G--I--V--L--D--S--G--D--G--V--T--H--T--V--P--I--Y--E--G--X--A--L--P--H--A--I--L--
701 GTCTCGACCTCGCGCGCGCACTCACCGACTCCTCATGAAGATCCTCACCGAGCGCGGTACTCCTTACCACCACCGCTGAGCGTGAATTCGTCGG
R--L--D--L--A--G--R--D--L--T--D--Y--L--M--K--I--L--T--E--R--G--Y--S--F--T--T--A--E--R--E--I--V--R
801 CGACATCAAGGAGAAGCTCTGCTACGTCCGCTGGACTTCGAGACCGAGATGCAGACCGCTCGACCTCCAGCACCCTCGAGAAGACCTACGAGCTTCGG
--D--I--K--E--K--L--C--Y--V--A--L--D--F--E--T--E--M--Q--T--A--S--T--S--S--T--V--E--K--T--Y--E--L--P
901 GACGGCCAGgtcagttatagataaaaaatgttctctctctcttttagacccccggcgccccgcagGTCATCACCATCGGCAACGAGGGCTTCCGGTG
-D--G--Q-----V--I--T--I--G--N--E--G--F--X--C
1001 CCCGAGGTCCTCTTCCAGCCAGCTTCATCGGCATGGAGTCCGGCCGCATCCAGGAGACCACCTACAACAGgtccggtttccttctgtgctgcctccc
--P--E--V--L--F--Q--P--S--F--I--G--M--E--S--A--G--I--H--E--T--T--Y--N--R-----
1101 tccgcttcgagctgacccggttcccgagCATCATGAAGTGCAGCTCGATATTCGTAAGACCTGTACGGCAACATCGTCTCCTCCGGCGCACCCACAT
-----I--M--K--C--D--V--D--I--R--K--D--L--Y--G--N--I--V--L--S--G--G--T--T--M
1201 GTACGCTGGCATCNGNACCCGATGAGAAAGGAGgtccgtagttatctggttttcggtctcctggttttagattttaaccctttgcccagATCACCCGCCCT
--Y--A--G--I--X--D--R--M--Q--K--E-----I--T--A--L
1301 GGGCCCCGTCGATGAAGATCAAGATCGTCCGCCCGGAGCGCAAGTACTCTGCTGGATCGCGGGgttcgccccttttctcgttatgtgtgtccaaa
--A--P--S--S--M--K--I--K--I--V--A--P--P--E--R--K--Y--S--V--W--I--G--G-----S--I--L--A--S--L--S--T--F--Q--M--W--I--S--K--Q--E--Y--D--E--S--G--P
1401 tcagcccactgaaagcgggttcgctcagGTCTATCCTTGCCTCGCTCCACCTTCCAGCAGATGTGGATCTCGAAGCAGGAGTACGACGAGTCCGGCCCC
-----S--I--L--A--S--L--S--T--F--Q--M--W--I--S--K--Q--E--Y--D--E--S--G--P
1501 TCGATCGTCCACAGGAATGCTTT*
-S--I--V--H--R--K--C--F
    
```



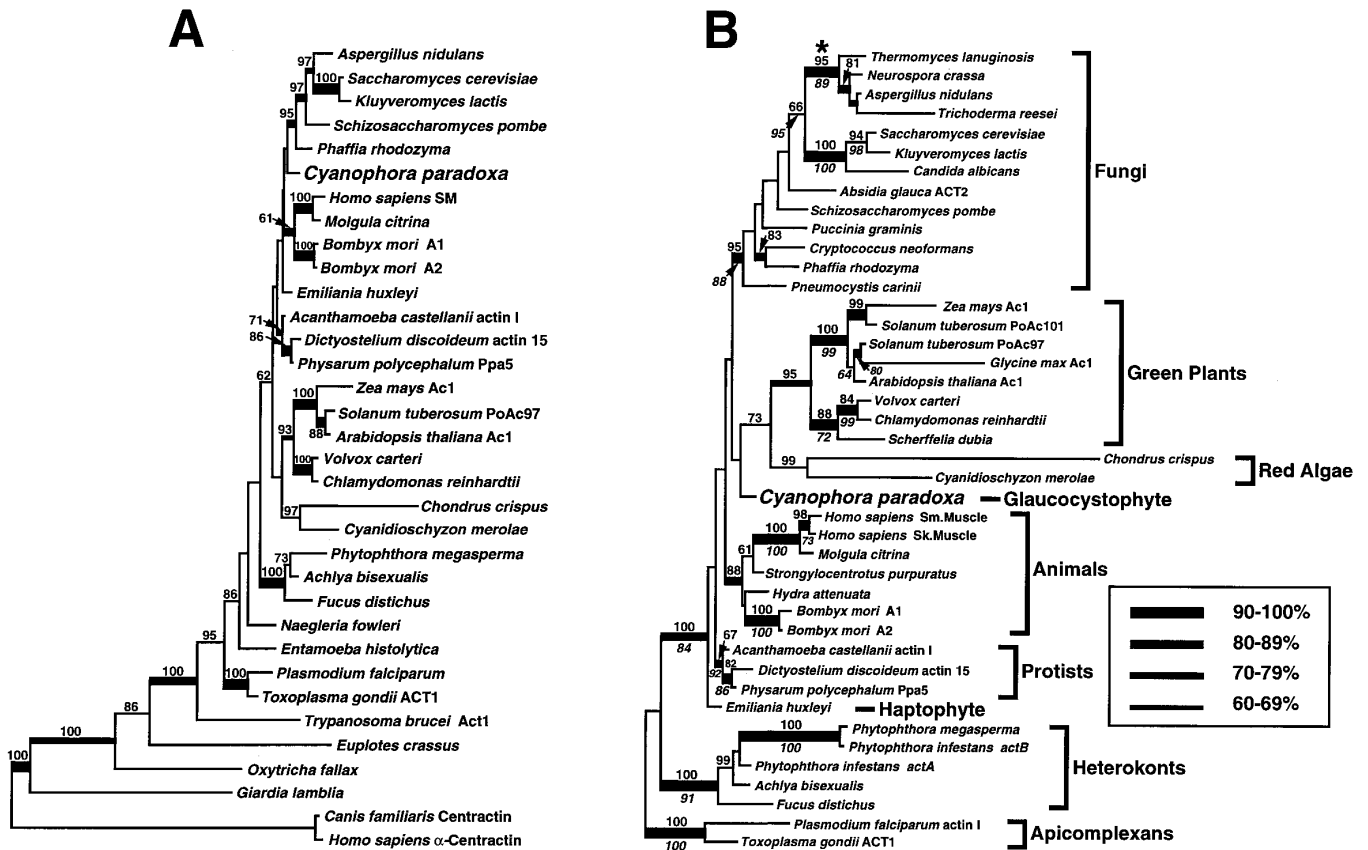


Fig. 3 A, B Phylogeny of actin-coding regions. **A** tree inferred with a weighted (RC index over the interval 1–1000) MP method (PAUP V3.1.1, Swofford 1993) using a heuristic search procedure with a branch-swapping algorithm (TBR, tree bisection-re-connection) and the first and second positions of actin codons (748 nucleotides). The bootstrap (100 replications)-consensus phylogeny shown has a consistency index of 0.620. The bootstrap values are presented at the internal nodes. This tree has been rooted within the branch-length leading to the centractin sequences. **B** phylogenetic analysis of actin coding regions in crown group eukaryotes done as described above. The bootstrap (200 replications)-consensus phylogeny shown has a consistency index of 0.686. The bootstrap values presented above the internal nodes were inferred from the MP analyses, whereas the values shown below the nodes were inferred from a bootstrap analysis (100 replications) using the actin protein sequences and the PROTDIST method (Phylip V.3.5c, Felsenstein 1993) with a neighbor-joining tree re-construction. This tree has been rooted within the branch length leading to the apicomplexan actin sequences. The results of bootstrapped distance analyses (200 replications) of both data sets using the first and second positions of the actin codons with a Kimura (1980) model and a neighbor-joining tree reconstruction are shown as branch lengths of differing thicknesses. The legend to these bootstrap values is shown next to **B**. The asterisk in **B** marks the putative origin of a homologous family of introns found in the actin genes of the ascomycetous fungi to the right of this symbol (see text for details). Only bootstrap values $\geq 60\%$ are shown and are indicated with arrows where they do not fit on the branches

was done using the primers 613F and DBAc3 with genomic DNA from *Cyanophora*. A fragment of approximate size 760 nt was isolated and verified as encoding actin using a Southern analysis with the *Cyanophora* cDNA fragment as a probe. These data suggest that no more introns inter-

rupt the 3'-terminal end since the expected size of the genomic actin sequence with the 3'-terminal sequence determined from the cDNA is also 760 nt. The *Cyanophora* actin introns all have the consensus $GT^C/T^A/C$ with CAG donor and acceptor sequences at the 5' and 3' termini of the introns, respectively (Fig. 2 B, Jackson 1991). A putative branch-point sequence was identified in each *Cyanophora* actin intron (underlined in Fig. 2). The consensus of these branch-point sequences (shown below Fig. 2) maintains an adenosine at the branch site that has been shown to be crucial for the chemical steps of splicing (Query et al. 1996). The *Cyanophora* consensus branch-point sequence is similar to that in higher plants (i.e. $CT^A/G^A/C^T$, Sinibaldi and Mettler 1992).

The *Cyanophora* coding region was readily aligned with other actin sequences and the phylogenetic analyses of the DNA and protein sequences position this alga as an independent lineage within the crown-group radiation (Fig. 3 A, B). Actin sequences are needed from cryptophyte algae to test the sister-group relationship between the glaucocystophytes and cryptophytes identified in the rDNA phylogenies (Bhattacharya and Medlin 1995). Inclusion of the centractin sequences as an outgroup identifies the archezoan *Giardia* as the earliest divergence within the eukaryote actin phylogeny (Fig. 3 A, see also Baldauf and Palmer 1993; Drouin et al. 1995). Many of the major crown-group lineages (e.g. fungi, animals, green plants, heterokonts) form monophyletic groups with bootstrap support in the actin phylogenies (see in particular Fig. 3 B).

The ciliate actin sequences (i.e. *Euplotes c.*, *Oxytricha f.*) do not however group either with themselves or with the apicomplexans, as would be expected on the basis of rDNA and morphological analyses (Cavalier-Smith 1993). This result may stem from the high divergence of the ciliate actin-coding regions and their resultant homoplasious "attraction" (Swofford and Olsen 1990) to the outgroup taxa (see also Drouin et al. 1995).

The fungi and animals which form a lineage with no bootstrap support in Fig. 3 A do not share a common origin in the the phylogeny shown in Fig. 3 B or in the protein distance tree (data not shown). This result is at odds with comparisons of rDNA (Wainright et al. 1993) and protein-coding (e.g., α -, β -tubulin, EF-1 α ; Baldauf and Palmer 1993) sequences and previously published actin phylogenies (Bhattacharya and Ehrling 1995; Drouin et al. 1995). A likely explanation for this inconsistency is the inclusion of actin sequences from protists that are either single representatives of their phylum or have very low sequence-divergence rates (e.g. slime molds) in the phylogenetic analyses (see Baldauf and Palmer 1993 for discussion). Exclusion of the *Acanthamoeba c.*, *Dictyostelium d.*, *Physarum p.*, *Emiliana h.* and *Cyanophora* actin sequences from the data set shown in Fig. 3 B resulted, for example, in bootstrap support for the monophyly of the fungi and animals. The distance analysis of nucleotides encoding the first and second positions of actin codons resulted in a bootstrap value of 85% (200 replications), while the MP analysis of the same data set resulted in a bootstrap value of 83% (100 replications) in support of the animal-fungal clade; the distance tree also grouped the green plants and red algae together as in the weighted MP analysis (Fig. 3 B) though with no bootstrap support (56%). Analysis of the reduced data set of protein sequences resulted in a bootstrap value of 80% in support of the animal-fungal clade.

An additional reason for the weak bootstrap support for some well-established monophyletic groups in the actin phylogenies (e.g. animals in the MP analysis shown in Fig. 3 A, 61%) is most likely due to variation in sequence-divergence rates. A high sequence-divergence-rate variation within and between lineages (Bhattacharya et al. 1991) may be the most important problem that hinders the use of actin and other molecular markers as gene/species markers (see also Palmer and Delwiche 1996). It has been shown, for example, with the relative-rate test that *Dictyostelium* and *Physarum*, which unexpectedly form a cluster within the crown-group in actin phylogenies (Fig. 3 B, Bhattacharya et al. 1991, 1993; Drouin et al. 1995), have a significant sequence-divergence-rate slowdown in comparison to other crown-group members (e.g. green plants, Drouin et al. 1995). *Dictyostelium* and *Physarum* diverge well before the crown-group radiation in rDNA sequence analyses (Sogin 1991; Cavalier-Smith 1993; McFadden et al. 1994).

In spite of these difficulties the use of protein coding-sequence phylogenies is an important component of molecular evolutionary studies. One result, for example, that was found in the MP analyses of both data sets, as well as

in the distance analyses of the second data set with and without *Acanthamoeba*, *Dictyostelium*, *Physarum*, *Emiliana* and the *Cyanophora* actin sequences, appears to be promising and deserves further study. These trees are consistent with a sister-group relationship between green plants and red algae and therefore with a monophyletic origin of the plastid in these groups. In plastid SSU rDNA and protein (except *rbcS,L*) phylogenies, the red- and green-algal plastids form monophyletic lineages that are sister groups of each other with a common root within the cyanobacteria. The glaucocystophyte cyanelles form a third plastid lineage that is also closely related to red- and green-algal plastids (Morden et al. 1992; Helmchen et al. 1995). The actin phylogenies suggest that the red- and green-algae may have had a single origin of their plastids in their common ancestor. The nuclear-encoded rDNA sequence comparisons have not yet been able to identify conclusively the sister group of green plants and red algae and have therefore not been useful for tracing the origin(s) of their plastids (Bhattacharya and Medlin 1995). As more actin sequences become available from these groups, it will be interesting to see if the glaucocystophytes also group with bootstrap support with the red algae and green plants (as in Fig. 3 B).

Actin intron catalogue

Four of the five *Cyanophora* actin introns (44-1, 283-2, 318-3, 345-2) occupy novel sites in the catalogue of actin intron positions, whereas the fifth intron (248-3) is also found in the actin gene of the green algae *Volvox c.* and *Chlamydomonas r.* and in two genes of the crustacean *Artemia* (Fig. 4). The sizeable increase in the total number of actin intron positions from 40 (Weber and Kabsch 1994; Sheterline et al. 1995) to 56 (Fig. 4) is due to the addition of four *Cyanophora* introns (see above), one *Artemia* intron (303-1), and the 11 fungal introns (3-2, 7-1, 10-3, 23-2, 34-2, 57-2, 121-1, 256-2, 312-1, 339-2, 359-1) found in *Phaffia r.*, *Cryptococcus n.* and *Puccinia g.* On the protein-sequence level, 2 of these 16 new intron positions (34-2 and 44-1) are not resolved from two previously documented introns (34-1 and 44-2) and one position (3-2) lies in the variable N-terminal region. Thus 13 new intron positions can be analyzed.

A previous analysis of 36 intron positions on the protein-sequence level showed that 23 positions are found at the ends of secondary structural elements (β strand, α -helix and 3_{10} helix) plus or minus one residue. Thirteen positions do not show this relationship and eight interrupt a secondary structural element. Statistical analysis showed that, due to the large number of end positions in the actin molecule, the boundaries of secondary structural elements are not correlated with intron positions (for details see Weber and Kabsch 1994). Using the same approach for the 13 novel intron positions reported here, eight are found at the ends of secondary structural elements plus or minus one residue (10-3, 23-2, 57-2, 256-2, 303-1, 312-1, 339-2 and 359-1) whereas five are found elsewhere (7-1, 121-1,

283-2, 318-3 and 345-2). Thus there is a random distribution of intron positions with regard to the secondary structural elements of actin (Weber and Kabsch 1994). Using several proteins of known three-dimensional structure and the corresponding gene organisations Stoltzfus et al. (1994) found no significant correspondence between exons and units of protein structure.

Evolutionary analyses of actin introns

The *Cyanophora* intron at position 44-1 is separated by only one or two nucleotides from two earlier-established intron positions (Fig. 4). Position 43-3 is relatively common in metazoa and has recently also been found as the sole intron in the actin gene of the red alga *Chondrus c.* (Bouget et al. 1995). An intron at position 44-2 is present in the single-copy actin genes of the fungi *Thermomyces l.*, *Aspergillus n.*, *Neurospora c.* and *Trichoderma r.* These three closely spaced intron positions (i.e. 43-3, 44-1, and 44-2) could either reflect independent insertions at a hot spot or they could have arisen by sliding of an ancient intron by one, or at most two, nucleotides (see Cerff 1996 for a review of intron sliding). Given the strict length conservation of the actin molecule, except for the very N-terminal end, it does not seem likely that intron sliding is accompanied by the loss or gain of one or more amino acid-residues. However, the minimal mutational events needed to slide an intron by one nucleotide can be a single-nucleotide insertion followed by a compensating deletion, or the reverse order of events. Although the first resulting gene product would be a truncated non-functional actin, the second mutational event could restore a normal actin molecule. Such a mechanism, involving two compensating events, could be tolerated in organisms with several actin

genes and possibly also in organisms with a single actin gene per haploid genome. Two additional hot spots that could also be examples of intron sliding are shown in Fig. 4. *Thermomyces* and three other fungi with identical actin intron patterns contain an intron at position 34-1 that is one nucleotide removed from the 34-2 intron of the fungi *Cryptococcus* and *Phaffia*. The *Thermomyces* fungal group also contains an intron at position 15-2, one nucleotide removed from the 15-1 intron of the *Drosophila* actin-4 gene. Since the latter position is unique among the various metazoan genes, including the five other *Drosophila* genes, we favor the view of an independent insertional event rather than intron sliding.

Based on the intron positions in triose-phosphate isomerase (TIM) genes, Kwiatowski et al. (1995) and Logsdon et al. (1995) argued against the exon theory of genes (Gilbert et al. 1986) and emphasized instead a random insertional origin of introns. Twelve of the twenty one positions within the TIM intron catalogue are thought to be of recent origin. This conclusion is most convincing for intron 5, which is found in only two mosquitos (*Culex* and *Aedes*, but not in a further mosquito *Anopheles*), four other insects, and all other TIM genes analyzed (Kwiatowski et al. 1995; Logsdon et al. 1995). A tentative assignment of actin intron positions in terms of evolutionary origins is indicated in Fig. 4. Introns marked as "new" are lineage- (e.g. metazoa, fungi or plants) or species-specific. Introns marked as "old" occur in at least two lineages and may have been present in the common ancestor of crown-group lineages. For none of the "new" introns in the actin catalogue is the assignment of proposed time of origin as clearly delineated as in the case of intron 5 of the TIM catalogue (see above). However actin intron 330-3 is indeed a good candidate for a "new" intron. Found in all vertebrate actin genes it is absent from the sister group of the urochordates (*Halocynthia*) and other deuterostomia (echinoderms; sea urchins and the sea star) as well as all other known genes. Thus it may be an intron position which was acquired within the vertebrates. The four introns of *Cyanophora* that occupy unique positions in the catalogue can be considered as "new" introns. We are aware, however, of the tentative nature of the assignment of a "new" intron in the actin catalogue. In the previous catalogue an intron at position 248-3, for example, seemed specific for the green alga *Volvox* (Weber and Kabsch 1994; Sheterline et al. 1995). Meanwhile it has not only been found in *Chlamydomonas* but also in the glaucocystophyte *Cyanophora* (this study) and in two genes of the arthropod *Artemia* (Fig. 4). Thus a putatively „young“ intron can readily turn into an „old“ intron once more genes become known. We also note some uncertainty about the assignment of "old" introns. An intron position which occurs in two lineages, but is only documented by one species, can also reflect two independent insertional events (Fig. 4). Some of these questions may be resolved once more actin genes become available. Alternatively, a more systematic approach to the analysis of actin intron positions in particular lineages or groups of organisms may provide more insights into the origin and loss of introns.

abbreviations 'm.' and 'cytoplasmic' stand for muscle and cytoplasmic actins, respectively. Sea urchin genes are from *Strongylocentrotus p.* and *Strongylocentrotus f.*, respectively. Sea star genes are from *Pisaster o.* Numbers following the species refer to the number(s) of the particular gene(s). Numbers in parenthesis indicate the total number of distinct genes sharing the particular intron pattern. The three-intron pattern of *Arabidopsis*, soybean, potato, tobacco, rice and maize is referred to as higher plants. The *Absidia g.* gene lacks information past position 140. An actin gene of *Ceratitis c.* (Mediterranean fruitfly) corresponds to *Drosophila* genes 1 and 6. Intron positions coinciding with the start or end of a secondary structural element of the actin molecule plus/minus one residue are marked by a plus sign in the top line of the figure. Introns unrelated to such a position are indicated by a minus sign. The first three intron positions are not considered since they occur in the only length-variable region of actin and are found only in some fungi (for details see Weber and Kabsch 1994). Tentative assignments of the "age" of actin introns are shown at the bottom of the figure. In this scheme, "new" introns are lineage- or species-specific, "old" introns are shared between at least two lineages and "slides" are introns that may trace their origin to a single intron that has been translocated 1 or 2 nucleotides from the original site. Circles mark "old" introns that could also be interpreted as representing independent insertional events in two lineages (see text for details)

Acknowledgements D. B. was supported by a grant from the Deutsche Forschungsgemeinschaft (Bh 4/1-2). We thank Wolfgang Berning-Koch for technical assistance and Juergen Ehling (University of Cologne) for construction of the *Cyanophora* genomic library.

References

- Baldauf SL, Palmer JD (1993) Animals and fungi are each others closest relatives: congruent evidence from multiple proteins. *Proc Natl Acad Sci USA* 90:11558-11562
- Bhattacharya D, Ehling J (1995) Actin-coding regions: gene family evolution and use as a phylogenetic marker. *Archiv Protistenkd* 145:155-164
- Bhattacharya D, Medlin L (1995) The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J Phycol* 31:489-498
- Bhattacharya D, Stickel SK, Sogin ML (1991) Molecular phylogenetic analysis of actin genic regions from *Achlya bisexualis* (Oomycota) and *Costaria costata* (Chromophyta). *J Mol Evol* 33:535-536
- Bhattacharya D, Stickel SK, Sogin ML (1993) Isolation and molecular phylogenetic analysis of actin-coding regions from the prymnesiophyte alga, *Emiliania huxleyi*, using reverse transcriptase and PCR methods. *Mol Biol Evol* 10:689-703
- Bhattacharya D, Helmchen T, Bibeau C, Melkonian M (1995) Comparisons of nuclear-encoded small-subunit ribosomal RNAs reveal the evolutionary position of the Glaucocystophyta. *Mol Biol Evol* 12:415-420
- Bouget F-Y, Kerbouch C, Liaud M-F, de Goër SL, Quatrano RS, Cerff R, Kloareg B (1995) Structural features and phylogeny of the actin gene of *Chondrus crispus* (Gigartinales, Rhodophyta). *Curr Genet* 28:164-172
- Cavalier-Smith T (1982) The origins of plastids. *Biol J Linn Soc* 17:289-306
- Cavalier-Smith T (1993) Kingdom protozoa and its 18 phyla. *Microbiol Rev* 57:953-994
- Cerff R (1996) The chimeric nature of nuclear genomes and the antiquity of introns as demonstrated by the GAPDH gene system. In: Go M, Schimmel P (eds) *Tracing biological evolution in protein and gene structures*. Elsevier, Tokyo, pp 205-227
- Clark, SW, Meyer DI (1992) Centractin is an actin homologue associated with the centrosome. *Nature* 359:246-250
- Dibb NJ, Newman AJ (1989) Evidence that introns arose at protosplice sites. *EMBO J* 8:2015-2021
- Drouin G, Moniz de Sà M, Zuker M (1995) The *Giardia lamblia* actin gene and the phylogeny of eukaryotes. *J Mol Evol* 41:841-849
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783-791
- Felsenstein J (1993) *PHYLIP Manual*, Version 3.5. Department of Genetics, University of Washington, Seattle
- Gilbert W, Marchionni M, McKnight G (1986) On the antiquity of introns. *Cell* 46:151-154
- Helmchen T, Bhattacharya D, Melkonian M (1995) Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J Mol Evol* 41:203-210
- Jackson IJ (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res* 19:3795-3798
- Kies L (1979) Zur systematischen Einordnung von *Cyanophora paradoxa*, *Gloeochaete wittrockiana* und *Glaucocystis nostochinearum*. *Ber Deutsch Bot Ges Bd* 92:445-454
- Kies L, Kremer BP (1990) Phylum Glaucocystophyta. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds) *Handbook of protocista*. Jones and Bartlett, Boston, pp 152-166
- Kimura, M (1980) A simple method for estimating evolutionary rates of base substitution through comparative studies of sequence evolution. *J Mol Evol* 16:111-120
- Kwiatowski J, Krawczyk M, Kornacki M, Bailey K, Ayala FJ (1995) Evidence against the exon theory of genes derived from the triose-phosphate isomerase gene. *Proc Natl Acad Sci USA* 92:8503-8506
- Lees-Miller JP, Helfman DM, Schroer TA (1992) A vertebrate actin-related protein is a component of a multisubunit complex involved in microtubule-based vesicle motility. *Nature* 359:244-246
- Logsdon Jr JM, Tyshenko MG, Dixon C, Jafari JD, Walker VK, Palmer JD (1995) Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. *Proc Natl Acad Sci USA* 92:8507-8511
- Loomis WF, Smith DW (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein-sequence comparisons. *Proc Natl Acad Sci USA* 87:9093-9097
- McFadden GI, Gilson PR, Hill DRA (1994) *Goniomonas* - rRNA sequences indicate that this phagotrophic flagellate is a close relative of the host component of cryptomonads. *Eur J Phycol* 29:29-32
- Melkonian M (1983) Evolution of green algae in relation to endosymbiosis. In: Schenk HEA, Schwemmler W (eds) *Endocytobiology II*. W. de Gruyter, Berlin, pp 1003-1007
- Moniz de Sà M, Drouin G (1996) Phylogeny and substitution rates of angiosperm actin genes. *Mol Biol Evol* 13:1198-1212
- Morden CW, Delwiche CF, Kuhse M, Palmer JD (1992) Gene phylogenies and the endosymbiotic origin of plastids. *BioSystems* 28:75-90
- Ortega M-A, Díaz-Guerra M, Sastre L (1996) Actin gene structures in two *Artemia* species, *A. franciscana* and *A. parthenogenetica*. *J Mol Evol* 43:224-235
- Palmer JD, Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proc Natl Acad Sci USA* 93:7432-7435
- Query CC, Strobel SA, Sharp PA (1996) Three recognition events at the branch-site adenine. *EMBO J* 15:1392-1402
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406-425
- Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467
- Schlösser UG (1984) Sammlung von Algenkulturen: additions to the collection since 1982. *Ber Deutsch Bot Ges* 97:465-475
- Sheterline P, Clayton J, Sparrow JC (1995) Actin. *Prot Profile* 2:27-28
- Sinibaldi R, Mettler I (1992) Intron splicing and intron-mediated enhanced expression in monocots. *Prog Nucleic Acid Res Mol Biol* 42:229-257
- Sogin ML (1991) The phylogenetic significance of sequence diversity and length variations in eukaryotic small-subunit ribosomal RNA coding regions. *Wistar Symp Ser* 4:175-188
- Sogin ML, Gunderson GJ, Elwood HJ, Alonso RA, Peattie DA (1989) Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. *Science* 243:75-77
- Stoltzfus A, Spencer DF, Zuker M, Logsdon JM, Doolittle WF (1994) Testing the exon theory of genes: the evidence from protein structure. *Science* 265:202-207
- Swofford DL (1993) *PAUP: Phylogenetic analysis using parsimony*, V3.1.1. Smithsonian Institution, Washington DC
- Swofford DL, Olsen GJ (1990) Phylogeny reconstruction. In: Hillis DM, Moritz C (eds) *Molecular systematics*. Sinauer, Sunderland, pp 411-501
- Wainwright PO, Hinkle G, Sogin ML, Stickel SK (1993) Monophyletic origins of the Metazoa: an evolutionary link with the fungi. *Science* 260:340-342
- Weber K, Kabsch W (1994) Intron positions in actin genes seem unrelated to the secondary structure of the protein. *EMBO J* 13:1280-1286