The Origin and Evolution of Green Algal and Plant Actins

Seon Sook An, Barbara Möpps, Klaus Weber, Debashish Bhattacharya

Introduction

Actin is one of the most highly conserved and best studied eukaryotic proteins. Actin shares a common evolutionary history with actin-related proteins (ARPys) (Clark and Meyer 1992; Lees-Miller, Helfman, and Schroeo 1992) and, in the more recent past, with other ATPases such as hexokinase and the 70-kDa heat-shock proteins common to all eukaryotes (Bork, Sander, and Valencia 1992; Sheterline, Clayton, and Sparrow 1995; Bhattacharya and Weber 1997). Actin generally occurs in complex families in multicellular organisms (e.g., animals, land plants) and as single copies in many protists (e.g., ciliates, fungi, red algae, most green algae, diplomonads [Bhattacharya and Ehling 1995; Drouin, Moniz de Sá, and Zuber 1995]). The role of actin in muscle contraction and its importance as a major component of the cytoskeleton has led to the detailed characterization of actin gene families in animals. Such data show tissue-specific expression of actins, supporting the hypothesis that the multiple actin genes encode proteins with specific and different cellular functions (see Herman 1993; Sheterline, Clayton, and Sparrow 1995). The angiosperms also contain complex and relatively highly divergent actin gene families (Hightower and Meagher 1986; Baird and Meagher 1987; McLean et al. 1990; Moniz de Sá and Drouin 1996). The relationship between the large number and subclasses of angiosperm actins and their function(s) is, however, unclear.

Key words: actin, evolution, gene duplication, green algae, plants, phylogeny.

Address for correspondence and reprints: Debasish Bhattacharya, Department of Biological Sciences, University of Iowa, 239 Biology Building, Iowa City, Iowa 52242-1324. E-mail: dbhattac@blue.weeg.uiowa.edu.


© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

There are at least three scenarios that could explain the existence of the complex actin gene families in angiosperms. Meagher and coworkers suggest, for example, that angiosperm actins are of an ancient origin (i.e., as old as animal actins [Hightower and Meagher 1986; Meagher 1991]) and that the increase in the complexity of this gene family has accompanied and may have played a direct role in the morphological development of flowering plants (Meagher 1995; McDowell et al. 1996). This hypothesis is supported by recent results on differential expression of actin genes in different tissues of Arabidopsis thaliana (An et al. 1996; Huang et al. 1996a; McDowell et al. 1996) and the general observation that actin is an important component of the plant cytoskeleton. Plant microfilaments control cell shape and are involved in plane localization during cell division. Actin has also been localized in many vegetative and reproductive structures and is postulated to be involved in organellar movement and cytoplasmic streaming (Williamson 1993).

A second scenario is that the evolution of actin genes in angiosperms may have less to do with the origin of complex tissues than with relaxed constraints on actin evolution during the diversification of these taxa. Moniz de Sá and Drouin (1996) have shown, in an extensive study of over 50 partial angiosperm actin-coding regions, that the high divergence of these sequences is due to relatively higher rates of sequence evolution rather than to an ancient origin. It is therefore conceivable that the isofrom number and sequence diversity of angiosperm actins reflect elevated rates of gene duplication and divergence in the genomes of these taxa relative to those of animals. The nonsynonymous rate of sequence evolution for angiosperm actins, for example, is 6–19 times higher than that for mammalian actin genes (Moniz de Sá and Drouin 1996).
A third scenario is that actin gene redundancy exists to produce large amounts of this important protein. According to this view, the expression of evolutionarily related actin genes may be controlled by a common regulatory (developmental or tissue-specific) element, such that these coding regions encode a pool of similar proteins to perform a single function (Huang et al. 1996a).

The origin and phylogeny of actin genes and their role in the evolution of the angiosperms are therefore largely unresolved. In addition, virtually all existing studies on land plant actin gene origin and evolution (e.g., Hightower and Meagher 1986; McDowell et al. 1996; Moniz de Sá and Drouin 1996) have focused on analyses of angiosperms and have not included enough members of earlier-diverging lineages within the Streptophyta (sensu Bremer 1985). The Streptophyta contain the angiosperms and the gymnosperms within a larger clade that includes the charophyte algae, bryophytes, and ferns (Kenrick and Crane 1997). The Streptophyta are the sister group to the Chlorophyta (sensu Sluiman 1985), and together they form the Viridiplantae (Cavaliier-Smith 1981). Molecular phylogenetic studies show that the origin of the Viridiplantae can be interpreted as a set of evolutionary “steps” from a single-celled, scaly, biflagellate ancestor that gave rise on the one side to the green algae of the Chlorophyta and on the other side to the Streptophyta (Melkonian and Surek 1995; Graham 1996; Huss and Kranz 1997; Bhattacharya et al. 1998). The prasinophyte algae form a paraphyletic group that diverges at the base of these two lineages (Pickett-Heaps 1996; Huss and Kranz 1997; Bhattacharya et al. 1998). Recent analyses of actin sequences identify the single-celled prasinophyte Mesostigma viride as the earliest divergence within the Streptophyta (Bhattacharya et al. 1998); rDNA sequence analyses are consistent with this result (Melkonian and Surek 1995).

Here, we characterize one partial actin genomic sequence from Cosmarium botrytis (Zygmenatiales), two partial actin genomic sequences from Selaginella apoda (Selaginellales), three complete actin cDNA sequences from Anemia phyllitidis (Polypodiales), and one partial actin cDNA sequence from Psilotum triquetrum (Psilotales). These sequences were included in phylogenetic analyses with actin sequences from the Chlorophyta and Streptophyta to gain insights into the origin and diversification of actin genes in this eukaryotic clade.

Materials and Methods
Isolation of Genomic Actin Sequences

Total genomic DNA was extracted from an actively growing culture of C. botrytis (Desmidiaeae, SAG 136.80 [Sammlung von Algenkulturen Goettingen; Schloesser 1994]) and from a sporophyte of S. apoda (Greenhouse, University of Iowa) with the plant DNeasy kit (Qiagen). Actin genes from these taxa were amplified with the PCR method using the plant-specific forward primer Fern5 (5’-CTTGTCTGYGACAATGGATCWW GAATGGT-3’) and the reverse primer Ac3 (Bhattacharya, Stickel, and Sogin 1993). The Fern5 primer recognizes a sequence just upstream of the first conserved intron in Streptophyta actin genes (20-3; i.e., after the 3rd nucleotide of the 20th amino acid). Actin genes in the Streptophyta contain three conserved introns at positions 20-3, 152-1, and 356-3 (Moniz de Sá and Drouin 1996). The Ac3-primer is complementary to the 3’ termini of actin genes. The PCR conditions were a pretreatment at 95°C for 10 min followed by 35 cycles of a denaturing step at 95°C for 1 min, an annealing step at 60°C for 2 min, and an extension step at 72°C for 4 min. A final extension step at 72°C for 10 min was included after the 35 cycles. The PCR analyses resulted in two distinct bands in the Selaginella amplification (genes Ac1, Ac2) and a single band in the Cosmarium amplification. These PCR products were purified with a Qiapquick PCR-purification kit (Qiagen) and sequenced over both strands using the PCR primers and other actin-specific oligonucleotides (Bhattacharya, Stickel, and Sogin 1991) with a fluorescent-labeled dideoxy terminator sequencing method on an ABI 373A sequencing device.

Isolation of Actin cDNAs

Total RNA was purified from a sporophyte of Anemia using a GIT-extraction buffer and poly-A mRNA isolated according to Sambrook, Fritsch, and Maniatis (1989). A partial actin-encoding cDNA of approximately 600 nt was isolated from this mRNA using the RT-PCR method, and primers were directed against conserved regions of Streptophyta actin genes between introns 1 and 2 and between introns 2 and 3 (224F [5’-GAYATGGAAAAGATCTGG-3’] and 850R [5’-TCYACRTCRCAYTTTCATKAT-3’], respectively [Bhattacharya, Stickel, and Sogin 1993]). The Anemia actin fragment was radioactively labeled with [32P]dCTP with the random-primed method (Pharmaclia) and was used to screen a pBluescript II cDNA library of this species (kindly provided by H. Maucher, Ulm, Germany) using standard colony hybridization methods (Sambrook, Fritsch, and Maniatis 1989). Three plasmids (Anemia Ac1, Ac2, Ac3) encoding different actin-coding regions were identified with this method and sequenced over both strands with the dideoxy termination method using T3, T7, and actin-specific oligonucleotides (Bhattacharya, Stickel, and Sogin 1991) with radioactively labeled [35S]dATP and the Sequenase 2.0 kit (USB).

Total RNA was prepared from Psilotum, and a partial actin cDNA was isolated using the RT-PCR method and the actin-specific PCR primers Ac1 and Ac3 (Bhattacharya, Stickel, and Sogin 1993). This PCR fragment, of approximately 1,100 nt, was cloned in the pBluescript II vector, and the insert in one of the recombinant plasmids (Psilotum Ac1) was sequenced over both strands.

Analyses of Actin Gene-Copy Number in Selaginella

Selaginella genomic DNA was treated with EcoRI, HindIII, and PstI endonucleases under standard conditions (NEB). The digested DNA was transferred onto a nylon membrane (Southern blot) and probed with an actin PCR product of 1,100 nt that had been isolated from Selaginella total RNA using the RT-PCR method and the Ac1 and Ac3 primers (Bhattacharya, Stickel,
and Sogin 1993). The Southern hybridization was done using a nonradioactive method according to the manufacturer's instructions (Gene Images, Amersham). The hybridization was done in 5 × SSC, 0.1% (w/v) SDS, and 5% (w/v) dextran sulphate at 60°C overnight. Filters were initially washed for 15 min at 60°C in 1 × SSC, 0.1% SDS and then for 15 min at the same temperature in 0.1 × SSC, 0.1% SDS prior to detection.

We also studied the number of actin genes in Selaginella using the PCR method. In this analysis, forward primers that recognize different conserved sequences close to the 5' termini of actin-coding regions were used in combination with the highly conserved Ac3 reverse primer and genomic DNA of Selaginella, Mesostigma, and Zamia (Cycadales). We postulated that the introns in the genomic actin sequences from these taxa would vary in length (or be lost) among members of the gene family, thereby allowing the resolution of the different genes as distinct fragments on agarose gels. The forward primers used for this analysis were Fern5 and two general actin primers (142F: 5'ATGGACCA-GAAGGACGC-3'; 244F: 5'-GAYATGGAAAGATCT GC-3' [see Bhattacharya, Stickel, and Sogin 1993]). The 142F and 244F primers recognized conserved sequences between the 20-3 and 152-1 introns. The PCR conditions were as described above.

Alignments

All complete Chlorophyta and Streptophyta actin sequences (36 genes) were aligned manually using the Seq-App program (Gilbert 1992), and first and second positions of actin codons (748 nt) were prepared for phylogenetic analyses (data set 1) using the SeqEdit program (Olsen 1990). Actin sequences in data set 1 and GenBank accession numbers are as follows: A. phyllitis (Ac1 [AF091808], Ac2 [AF091809], Ac3 [AF091810]); A. thaliana (Act1 [M20016], Act2 [U37281], Act3 [U39480], Act4 [U27980], Act7 [U27811], Act8 [U42007], Act11 [U27891], Act12 [U27892], Clamhomonas reinhardtii (DS0839); Coleochaete scutata (AF061019); Daucus carota (Act1 [X17526], Act2 [X17525]); Glycine max (Act1 [J01298], Act3 [J01297]); Mesostigma viride (AF061020); Nicotiana tabacum (Ac25 [X63603]; Oryza sativa (Ac1 [X15865], Ac2 [X15864], Ac3 [X15862], Ac7 [X15863]); Pisum sativum (Ac1 [X68649], Ac2 [X67660]); Scherffelia dubia (AF061018); Solanum tuberosum (Ac58 [X55749], Ac71 [X55750], Act5 [X55753], Ac97 [X55751], Ac101 [X55752]); Sorghum vulgar (Act1 [X79378]); Striga asiatica (Ac1 [U68461], Ac2 [U68462]); Volvoc carteri (M33963); and Zea mays (Ac1 [J01238]).

We excluded the third codon positions of actin genes from our phylogenetic analyses because these coding regions in the angiosperms have extreme nucleotide base composition bias (Bhattacharya, Stickel, and Sogin 1991; Moniz de Sá and Drouin 1996). Actin genes in monocots, for example, have a significantly lower (46.4 ± 1.7%) average GC content than do actin genes in dicots (53.7 ± 3.5%; Moniz de Sá and Drouin 1996). There appears, however, to be no significant correlation between the GC content of the first (R^2 = 0.091) and second (R^2 = 0.038) bases with that of the overall GC content of angiosperm actin genes. Third positions are strongly correlated (R^2 = 0.986) with actin sequence GC content (Moniz de Sá and Drouin 1996). Exclusion of third codon positions from the phylogenetic analyses was expected to prevent the grouping of actin sequences in trees on the basis of shared nucleotide content rather than common ancestry (Bhattacharya and Ehlting 1995; Drouin, Moniz de Sá, and Zuker 1995). To test this line of reasoning, we prepared another data set (data set 2) for phylogenetic analyses that contained all of the sequences in data set 1 but included the third codon positions (1,128 nt).

A third data set (data set 3) was created from all available full-length and partial actin sequences from the Chlorophyta and Streptophyta, including the actin-coding regions of Selaginella, Cosmarium, Anemia, and Psilotum (see also the alignment in Moniz de Sá and Drouin 1996). This data set contained 86 sequences. A total of 688 nucleotides encoding first and second positions of actin codons were prepared for phylogenetic analyses from data set 3. The actin sequences in data set 3 (excluding those in data set 1) and GenBank accession numbers are as follows: Chlorella vulgaris (U66585); Cyanophora paradoxa (full-length cDNA, U90325); C. botryis (AF090970); Cucul volvulata (Cyc3 [AF002686], Cyc4 [AF002687], Cyc5 [AF002688]); G. max (Ac57 [U60500], Ac58 [U60499], Ac69 [U60498], Ac70 [U60497], Ac86 [U60496], Ac109 [U60506], Ac115 [U60505], Ac118 [U60503], Ac119 [U60502]); Lycoperniss esculentum (Ac32 [U60479], Ac41 [U60480], Ac51 [U60481], Ac52 [U60482], Ac105 [U60478]); N. tabacum (Ac53 [U60493], Ac54 [U60492], Ac66 [U60491], Ac71 [U60490], Ac93 [U60489], Ac103 [U60495], Ac104 [U60494]); Osmunda cinnamomea (Os14 [AF002685]); Podocarpus macrophyllus (Pod2 [AF002689], Pod4 [AF002690], Pod7 [AF002691]); P. triquetrum (Ac1 [AF091811]); S. apoda (Ac1 [AF090968], Ac2 [AF090969]); S. tuberosum (Ac42 [U60488], Ac46 [U60487], Ac65 [U60486], Ac79 [U60484], Ac82 [U60483]); Spirogyra sp. (AF061021); and Z. mays (Ac56 [U60514], Ac63 [U60513], Ac81 [U60511], Ac83 [U60510], Ac87 [U60509], Ac89 [U60508], Ac95 [U60507]). The actin sequences of Gluacosycystis nostochinearum, Microthamnion ketzianum, and Pseudocendolonium basileianse are unpublished and were kindly provided by T. Friedl, Goettingen, Germany. All three data sets used in the phylogenetic analyses have been deposited in the TreeBASE phylogenetic database (http://herbaria.harvard.edu/treebase; data set ID numbers are SN47-176, SN47-178, and SN47-179).

Phylogenetic Analyses

Data set 1 was subjected to maximum-likelihood analysis and to bootstrap (Felsenstein 1985) analyses using distance, LogDet, and maximum-parsimony using the PAUP* computer program (version 4.0b1; Swoford 1998). In the maximum-likelihood analysis, all sites were assumed to diverge at the same rate (HKY model;
Hasegawa, Kishino, and Yano (1985), the transition/transversion ratio was set to 2, the base frequencies were empirically determined, and the starting branch lengths were determined with the Rogers-Swofford approximation method (Swoford 1998). Phylogenies were built stepwise and rearranged using the tree-bisection-reconnection (TBR) branch-swapping algorithm to search for trees with a higher likelihood. In the distance bootstrap analysis (2,000 replications), the neighbor-joining method was used to build phylogenies from distance matrices calculated according to the HKY model. Neighbor-joining trees were also built from the matrices inferred with the LogDet method (2,000 bootstrap replications). In the maximum-parsimony bootstrap analysis (2,000 replications), sequence positions (154 parsimony-informative characters) were reweighted (using the rescaled consistency index over an interval of 1–1,000) to reduce the influence of highly divergent sites in the phylogeny reconstruction (Bhattacharya 1996). Starting tree(s) were obtained with stepwise (simple) addition and rearranged with TBR. The resulting bootstrap values from the neighbor-joining, LogDet, and maximum-parsimony analyses were included along the branches of the maximum-likelihood tree (for shared monophyletic groups). The phylogenies inferred from data set 1 were rooted with the actin sequence of _S. dubia_ (Prasinophyceae). Bootstrap analyses (500 replications) using the neighbor-joining method and matrices calculated with the LogDet and HKY models that contained all three actin codon positions were also done with data set 2. The LogDet method is applicable even when sequences vary in nucleotide content (Lockhart et al. 1994).

Data set 3 was subjected to bootstrap analyses (500 replications) using distance and LogDet methods as described above. The phylogenies inferred from data set 3 were rooted with the actin sequences of the glaucocystophyte algae, _C. paradoxa_ and _Glaucocystis_. The glaucocystophytes are photosynthetic protists that form a clade within the eukaryotic “crown group” (Helmchen, Bhattacharya, and Melkonian 1995).

**Results and Discussion**

**Gene Copy Number in the Chlorophyta and Streptophyta**

The determination of the copy number of actin genes in the Chlorophyta and the earlier-diverging Streptophyta is an important step in understanding the origin of these sequences in the land plants. The Chlorophyta are divided into the following classes: Chlorophyceae, Prasinophyceae, Trebouxiophyceae, and Ulvophyceae (Sluiman 1985). A detailed account of the phylogeny and evolution of these taxa can be found in Melkonian and Surek (1995) and Friedl (1997). Previous Southern and PCR analyses have suggested that two single-celled members of the Prasinophyceae, _Mesostigma_ and _Scherffelia_, and two members of the Chlorophyceae, _Chlamydomonas_ (single-celled) and _Volvox_ (colonial), contain a single actin gene (Cresnar et al. 1990; Bhattacharya et al. 1998). _Mesostigma_ is the earliest divergence within the Streptophyta, whereas _Scherffelia_ (Chlorodendrales) diverges within the near-simultaneous radiation of the “advanced” green algae (i.e., Chlorophyceae, Trebouxiophyceae, and Ulvophyceae; Stein-kötter et al. 1994; Surek et al. 1994; Friedl 1997). Presently, the only members of the Chlorophyta that are known to contain multiple copies of actin genes are two members of the Ulvophyceae, _Acetabularia cliftonii_ (GenBank accession number Z28698) and _Cladophora rupestris_ (T. Friedl, personal communication). The ulvophytes appear to have undergone independent actin gene duplications after their divergence from the other green algae.

In the present study, Southern analysis of _Selaginella_ genomic DNA resulted in the identification of two actin-anealing bands in the HindIII and _PstI_ digests (fig. 1A). In both of these digests, the larger fragment cross-reacted more strongly with the _Selaginella_ actin cDNA probe than did the smaller fragment. Only one strongly cross-reacting fragment was resolved in the EcoRI digest of _Selaginella_ DNA. These results are ambiguous regarding gene copy number in this taxon. There are presumably either one or two genes. If the latter, then the genes may be tandemly repeated at one locus, and the single strong bands represent both (or additional) copies. To gain additional insights into this problem, we used the PCR method with genomic DNA from _Selaginella_ and the conserved forward primers, _Fern5_ and _142F_, in combination with the conserved reverse primer Ac3 (Bhattacharya, Stickel, and Sogin 1993). These PCR reactions resulted in the amplification of two distinct actin-encoding fragments (see fig. 1B and below). PCR analyses of genomic DNA from _Mesostigma_ predictably showed single bands in all reactions, whereas, minimally, seven bands were identified in the _Zamia_ amplification. Gymnosperms contain multiple copies of actin genes (e.g., _C. revoluta_; Moniz de Sá and Drouin 1996 [GenBank AF002686–AF002688]). These results are consistent with there being (minimally) two actin genes in _Selaginella_. We realize, however, that the Southern and PCR analyses are not proof of actin
Evolution of Plant Actins

FIG. 2.—Phylogenetic analyses of the first and second codon positions of actin sequences (748 nt, data set 1) using PAUP* (version 4.0b1; Swofford 1998); tree inferred with the maximum likelihood method and the HKY model. The bootstrap values shown to the left of the commas above the branches were calculated with the LogDet method (2,000 replications), whereas the bootstrap values shown to the right of the commas were inferred from a weighted maximum-parsimony analysis (2,000 replications) using a stepwise (simple) sequence addition in a heuristic tree search. The bootstrap parsimony trees were rearranged with TBR. The bootstrap values shown below the branches in italic type were calculated with a distance method (HKY model, 2,000 replications). Both LogDet and distance bootstrap trees were built with the neighbor-joining method. Only bootstrap values ≥60% are shown. The vegetative (VEG) and reproductive (REP) classes of actin genes are shown in brackets.

Given these results, we postulate that the common ancestor of the Chlorophyta and Streptophyta contained a single, constitutively expressed, actin gene. Single-celled taxa such as Mesostigma and Scherffelia appear to have retained the single gene, whereas independent duplications of actin genes have occurred in the ulvophyte green algae and in Selaginella. Subsequent duplications have resulted in the complex actin families presently found in later-diverging land plants such as the ferns, cycads (see below), and the angiosperms (Moniz de Sá and Drouin 1996).

Phylogeny of Actin Genes in the Streptophyta

Maximum-likelihood analysis of the complete actin-coding regions (data set 1) recovered the tree shown in figure 2. This tree shows that the three Anemia actin sequences form a monophyletic group that diverges before the massive duplications of the angiosperm actin genes. Bootstrap analyses with the distance, LogDet, and maximum-parsimony methods support this result. Our data also suggest that the gene duplications within Anemia occurred independent of those in the angiosperms. Within the angiosperms, the actin sequences form two groups, vegetative (VEG) and reproductive (REP). These designations are based on limited expres-
The more extensive distance and LogDet analyses of all available full-length and partial actin coding regions of the Chlorophyta and the Streptophyta (data set 3), with the glaucocystophyte sequences as the outgroup, are shown in figure 3. This phylogeny has a topology similar to that resulting from the analysis of the full-length actin-coding regions. The three vegetative actin genes in Arabidopsis (ACT2, ACT7, and ACT8; McDowell et al. 1996) form a monophyletic group in figure 3, but without bootstrap support of 60%. These genes may have related functions, because they are the products of recent gene duplications. This same observation can be made for four of the five reproductive actin genes from Arabidopsis (ACT1, ACT3, ACT4, and ACT12; An et al. 1996; Huang et al. 1996a) that also form a monophyletic group. The astounding number of gene duplications in clusters of actin genes in some taxa (e.g., Z. mays Ac56, Ac63, Ac89) suggests either that the regulation of actin function is very complex in plants or that a great deal of redundancy occurs in this gene family. The LogDet and distance analyses also suggest that the REP actin genes originate from a paraphyletic group of VEG sequences. Although we did not find bootstrap support for a monophyletic origin of the REP genes, it is noteworthy that all actin genes that were in the REP class in figure 2 remain in this cluster after the addition of all available partial angiosperm actin sequences (taxon names capitalized in fig. 3). We interpret this result as support for a monophyletic derived group of angiosperm actin genes that are primarily expressed in reproductive tissue (in agreement with the results of McDowell et al. 1996).

Regarding the evolution of actin genes in the green algae and earlier-diverging land plants, many of the evolutionary relationships shown in figure 3 are consistent with accepted ideas of evolution in these groups. The actin genes from the Ulvophyceae and the Chlorophyceae form monophyletic groups, the Prasinophyceae are deeply paraphyletic (Mattox and Stewart 1984; Friedl 1997), the Charophyceae are early divergences within the Streptophyta, and the Selaginellina, Psilotum, Anemia, and Osmunda actin genes diverge prior to the major radiation of the angiosperms (Bhattacharya and Medlin 1998). This phylogeny also provides interesting insights into actin gene duplications. The Selaginella Ac1, Psilotum, Anemia, and Osmunda actin sequences form a monophyletic group that includes the Cycas Cyc3 gene. There are, however, no angiosperm actin genes within this group. These limited data suggest that actin gene duplications occurred in the common ancestor of the ferns and gymnosperms and that the angiosperm sequences are derived from particular members of these gene families. Interestingly, none of the gymnosperm actin genes branch within the angiosperm radiation (although these nodes lack bootstrap support 60%). This pattern agrees with distinct and separate evolutionary histories for gymnosperms and angiosperms (see Chaw et al. 1997), although the sampling here is limited.

The distribution of spliceosomal introns within the Cosmarium and Selaginella actin sequences is consistent with the results of the phylogenetic analyses. Both of
these genera contain the three introns (20-3, 152-1, 356-3) typical of Streptophyta actin genes (Moniz de Sá and Drouin 1996; Drouin and Moniz de Sá 1997). In addition, Cosmarium contains a novel intron at position 76-1 not yet found in any other Streptophyta or Chlorophyta. Two actin genes from the Chlorophyta (C. reinhardtii, V. carteri) are interrupted by eight introns, of which only the 20-3 intron is shared with the Streptophyta (see Bhattacharya et al. 1998 for details). Mesostigma contains the 20-3 and 152-1 introns but lacks the intron at position 356-3. Given the present data set, we postulate that the 20-3 intron may be the “oldest” intron in Streptophyta actin genes and that it may have existed in the actin gene of the common ancestor of the Chlorophyta and Streptophyta. The 356-3 intron apparently originated after the divergence of Mesostigma.
**Analysis of N-Terminal Actin Sequences**

The classification of multiple-copy actin genes, using N-terminal signature sequences that are conserved among taxa, is well established within the metazoa (Vandekerckhove and Weber 1984; Alonso 1987; Sheterline, Clayton, and Sparrow 1995). These studies show that the N-terminus of actins, which is normally acidic, varies greatly between actin isoforms within a given species and contains amino acid residues that are believed to be involved in actin function (e.g., myosin activation). The N-terminal end of animal actins is also an important binding site for several actin-binding proteins (for review, see Sheterline, Clayton, and Sparrow 1995).

We classified actins from the Chlorophyta and Streptophyta using conserved sequences found within the N-terminal residues (25 amino acids) of these proteins (fig. 4) and compared these data with the phylogeny of the complete actin-coding regions (fig. 2) to reach conclusions about the origin of the multiple-copy actin genes of the Streptophyta. The human $\beta$-non-muscle actin was used as the standard for these sequence comparisons to be consistent with the larger alignment published in Sheterline, Clayton, and Sparrow (1995).

The comparison of the N-termini of actins from the Chlorophyta and the Streptophyta reveals patterns that are consistent with the phylogeny of these sequences.
Proline-7, for example, is found in all vascular plant actins and is thought to play an important role in actin polymerization, because the substitution of this amino acid into animal actin and subsequent expression in yeast results in a 10-fold lower critical concentration for polymerization (Al-Almi et al. 1993). This apparently crucial amino acid substitution has occurred within the fern lineage (i.e., *Anemia*), given that the Chlorophyceae and the Charophyceae contain an alanine (A) at this position that is conserved in virtually all eukaryotic actins (except the ciliate *Oxytricha* spp. [Shetlerline, Clayton, and Sparrow 1995]). Furthermore, the residues aspartic acid (D), isoleucine (I), and glutamine (Q), which occur at positions 4, 5, and 6, respectively, of virtually all angiosperm actins, are found within the *Anemia* actin gene family. *Anemia* Ac3, for example, has aspartic acid-4 and isoleucine-5 (see fig. 4), whereas the *Anemia* Ac1, Ac2, and green algal actins encode glutamic acid (E) and valine (V) at these positions. These data suggest that the *Anemia* Ac3 gene is more closely related to the derived angiosperm actins than are either the Ac1 or the Ac2 genes from this taxon. The existence of a number of other amino acids (e.g., serine-14, and see below) shows, however, that the Ac3 gene also shares some ancestral features with green algal actins.

To identify other conserved residues that discriminate between the VEG, REP, and algal/fern actin genes, the phylogeny shown in figure 2 was used as input in MacClade (version 3.07; Maddison and Maddison 1997) with the aligned actin amino acid data set. Amino acids were identified that characterize these three types of actin genes: threonine-160 and serine-360 for the fern/algal actins, and serine-358 and valine-219 for the VEG and REP actin classes, respectively. Several amino acids are specific for all green algae (e.g., isoleucine-212) or for only the Chlorophyta (valine-272, phenylalanine-279, asparagine-295). The combination of the phylogenetic and the N-terminal sequence analyses allows the identification of amino acids that likely play an important role in actin function. Such residues as those that characterize the VEG and REP actin classes would be ideal targets for mutagenesis analyses to understand the role of these amino acids in actin function in the cell. In addition, these analyses show that the highly divergent angiosperm actins are the result of a stepwise evolutionary process that can still be reconstructed with the inclusion of green algal and fern sequences. Actin sequences such as those from *Mesostigma* or *Volvox* are descendants of the conserved single-copy actin gene that was found in the single-celled ancestor of all Chlorophyta and Streptophyta and was constitutively expressed in this organism. Residues such as alanine-7, serine-14, and isoleucine-212 were likely encoded in this ancestral actin gene, given that these residues are also found in the human β-non-muscle actin and in this protein from virtually all other eukaryotes (see the alignment in Shetlerline, Clayton, and Sparrow 1995).

Relationship Between the Phylogenetic Relationships of Angiosperm Actins and Their Functions

Do angiosperm actin genes with related functions also group together in phylogenies? If so, then the origin of actin genes may also have played a role in the origin of new actin functions in angiosperms. To address this issue, the known angiosperm actin mRNA localization data were mapped onto the phylogenetic tree shown in figure 2. The only taxon for which extensive localization data exist is *Arabidopsis* (An et al. 1996; Huang et al. 1996a; McDowell et al. 1996). Analyses with GUS reporter gene constructs and RNA hybridizations suggest that two basic types of actins are encoded in the *Arabidopsis* genome. One type is localized primarily in mature pollen and all organ primordia (*ACT1*, *ACT3*), in mature pollen and young vascular tissue (**ACT4**, **ACT12**), or in pollen, ovules, and developing embryos (**ACT11**, reproductive actin class of McDowell et al. 1996, REP class). The second type, encoded by the genes for **ACT2**, **ACT8**, and **ACT7**, is strongly expressed in vegetative structures but weakly or not at all in pollen (vegetative actin class of McDowell et al. [1996], VEG class). The limited expression data that exist for other angiosperm actins have also been mapped onto this tree. The *N. tabacum* Ac25 gene transcript, for example, is also found only in pollen RNA (Thangavelu et al. 1993) and is positioned within the REP class. Actin mRNA from the *Glycine* Ac3 gene is localized in leaves and is positioned as a sister group to the *Arabidopsis* vegetative actins in figure 2.

The separation of angiosperm actin genes into two clearly demarcated functional groups is, however, not possible because a large amount of overlap exists between the expression patterns of the different *Arabidopsis* genes (see McDowell et al. 1996) and because the "reproductive" group also contains members that are constitutively expressed (e.g., *Oryza* Ac1; McElroy et al. 1990). In fact, the two *Oryza* genes that are known to be constitutively expressed (McElroy et al. 1990) are found in both VEG (**AC7**) and REP (**Ac1**) actin classes, whereas genes that are largely expressed in pollen are found only in the REP class. The existence of constitutively expressed actin genes in both the VEG and the REP classes suggests that the ancestors of the angiosperm actin genes were also likely constitutively expressed. Further support for this hypothesis comes from northern analyses, with gene-specific flanking-sequence probes of the *Anemia* Ac1, Ac2, and Ac3 transcripts, which show that these mRNAs are found in all developmental stages (spore, gametophyte, sporophyte) and different tissues (trophophyll and sporophyll; B. Möpps, unpublished data). Actin localization data from *Chlamydomonas* also show that the single-copy actin gene in this taxon has many functions in the cell, including roles in cell division (Harper et al. 1992) and mating-tube formation (Detmers, Carboni, and Condeelis 1985).

Taken together, the gene expression and phylogenetic data are supportive of a scenario in which the specialized REP class of angiosperm actins has evolved from a relatively more primitive group of actins (VEG class; see fig. 3) which were constitutively expressed in vegetative structures. This is consistent with the idea that actin gene diversification in plants reflects the origin of new functions and/or tissues. In our analyses, the VEG actins share a common ancestry with fern genes.
that also appear to be constitutively expressed. That the multiple-copy fern actin genes all branch outside of the angiosperm clade also suggests that if actin gene origin and tissue origin are directly related, then morphological development within ferns has occurred independently of that in the angiosperms. Further support for these hypotheses will come from more extensive phylogenetic and expression studies with fern and other angiosperm actins (in progress). It would be interesting, for example, to determine whether pollen-specific actin mRNAs are limited to the REP class or whether this function has evolved multiple times within the angiosperms. The large number of actin isoforms in some plants (e.g., 24 ± 12 genes in *S. tuberosum* [Moniz de Sá and Drouin 1996], 25–30 genes in *Nicotiana* [Thangavelu et al. 1993]) suggests that there may be great overlap and/or duplication in actin function, and the regulation of this expression must therefore also be quite complex (see McDowell et al. [1996] for discussion). Preliminary analyses with the actin-binding protein (ABP) profilin from *Arabidopsis* show that this gene family also comprises distinct clades that are expressed in vegetative or reproductive structures (Huang et al. 1996b). This result raises the intriguing possibility that actin and ABPs may have coevolved during the evolution of the angiosperms, with the origins of different actin genes and their functions paralleling the diversification of ABPs such as profilin.

Duplication may not, however, be the only source of novel genes in land plants. Recent data show that polyploidization is a common (and recurrent) event in mosses, ferns, and angiosperms (see Soltis and Soltis 1995 for review). In the case of allotetraploid species such as *N. tabacum*, the redundancy in actin genes likely reflects both gene duplication events and the presumed doubling of acting gene copies through polyploidization. The increase in the number of distinct acting gene family members by polyploidization is also likely driven by gene diversification, resulting in the functional divergence of duplicated gene copies (Soltis and Soltis 1993). Our study does not address the role of polyploidization in the origin of land plant acting genes, nor do we attempt to distinguish between genes and alleles in the phylogenetic analyses. The level of divergence between most of the actin genes used in our study suggests, however, that these sequences are distinct gene family members. Possible exceptions are the two *P. sativum* (Ac1, Ac2) actins that differ at only two sites in their protein sequence. We expect that future studies that address the phylogeny and evolution of genes and genomes after polyploidization (e.g., Song et al. 1995) will provide a more complete understanding of the evolutionary history of actins in the Streptophyta.

Acknowledgments

This work was supported by generous funding from the College of Liberal Arts (University of Iowa) to D.B. We thank Anja Besendahl (Iowa) for technical assistance, Thomas Friedl and Alexandra Nickles for providing unpublished actin sequences, and two anonymous reviewers for their critical reading of this manuscript.

LITERATURE CITED


Pamela Soltis, reviewing editor

Accepted November 5, 1998