J Mol Evol (1997) 45:509–513

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1997

Loss of Introns in the Pollen-Specific Actin Gene Subfamily Members of Potato and Tomato

Guy Drouin, Mário Moniz de Sá*

Biology Department, University of Ottawa, 30 Marie Curie, Ottawa, ON, K1N 6N5, Canada

Received: 12 November 1996 / Accepted: 27 May 1997

Abstract. Most angiosperm actin genes, including the pollen-specific tobacco actin gene Tac25, contain three introns at conserved locations within their protein coding region. We have isolated one tomato and two potato actin genes that are related to this pollen-specific tobacco actin gene and which have lost one or two introns. The loss of these introns was likely the result of at least two gene conversion events with cDNAs derived from transcripts of pollen-specific genes.

Key words: Intron-loss events — Gene conversion — Reverse-transcribed cDNA — Actin — Pollen-specific — Potato — Tomato — Tobacco

Introduction

The genome of higher vertebrate species usually contains six different actin genes. Four of these genes code for muscle isoforms (α-skeletal, α-cardiac, α-vascular, and γ-enteric) and the two other genes code for the β- and γ-cytoplasmic isoforms (Kusakabe et al. 1997, and references therein). The positions of intron are conserved between orthologous vertebrate actin genes. For example, the coding region of the β-cytoplasmic gene in carp, *Xenopus,* chicken, rat, mouse, and human in each case has four introns at amino acid positions 41/42, 121/122, 267, and 327/328 (Carroll et al. 1986; Liu et al. 1990; Weber and Kabsch 1994; Bhattacharya and Weber 1997). These intron positions have therefore been conserved since the divergence of the lineages leading to fish and mammals some 400 MYA (Carroll 1988).

More than 60 angiosperm actin genes have so far been cloned and sequenced (reviewed in McDowell et al. 1996; Moniz de Sá and Drouin 1996). These include genes from dicotyledon species (potato, tomato, tobacco, *Arabidopsis thaliana,* soybean, pea, and carrot) as well as genes from monocotyledon species (maize, rice, and *Sorghum vulgare*). Except for the actin processed pseudogene found in potato, and the ACT2 gene of *A. thaliana* discussed below, all these genes contain three introns at amino acid positions 20/21, 152, and 355/356 (Drouin and Dover 1987; Sheterline and Sparrow 1994; McDowell et al. 1996; Moniz de Sá and Drouin 1996). These intron positions are also present in three actin genes from a conifer (*Podocarpus macrophalus*), three actin genes from a cycad (*Cycas revoluta*), and one actin gene from a fern (*Osmunda cinnamomea;* Moniz de Sá and Drouin, submitted). Conifers, cycads, and ferns are thought to have originated over 200, 225, and 280 MYA, respectively (Stewart 1983). The three actin gene introns found in the coding region of land plants have therefore been conserved for over 280 MYR.

McDowell et al. (1996) have recently reported the sequences of the 10 actin gene sequences present in the *A. thaliana* genome. One of these genes, ACT2, has had its first intron precisely deleted. They suggested that this intron was lost through a gene conversion event with an ACT2 cDNA. The fact that this gene has been shown to be the most highly expressed actin gene in *A. thaliana,* as

* *Present address:* Biotechnology Laboratory, University of British Columbia, 237-6174 University Boulevard, Vancouver, BC, V6T 1Z3, Canada
Correspondence to: G. Drouin; e-mail guy@bio01.bio.uottawa.ca

well as being expressed in tissues leading to reproductive organs and germ cells, is consistent with this suggestion (An et al. 1996).

Here we report the sequence of one tomato gene that has lost its second intron and two potato actin genes that have lost their second and third introns. We show that these genes are related to a pollen-specific tobacco actin gene and suggest that the intron loss events in the tomato and potato actin genes were also likely the result of at least two gene conversion events with cDNAs derived from actin genes expressed in the germ line. These results also indicate that intron loss events occur only in genomic sequences sharing high sequence similarity with the cDNA sequences derived from actin genes expressed in the germ line.

Materials and Methods

DNA Sequences. The actin genes Tom32, Pot42, and Pot65 were cloned using the polymerase chain reaction and sequenced as described previously (Moniz de Sá and Drouin 1996). The sequence of these genes were deposited in GenBank under the accession numbers U60479, U60488, and U60486, respectively. The Tac25 tobacco actin sequence was obtained from Genbank (accession number X63603). Manual alignment of the DNA sequences, based on the protein alignment, of exons 2 and 3 of these genes was straightforward due to the highly conserved nature of actin.

Relative Rate Tests. Synonymous and nonsynonymous substitutions were calculated using the method of Li et al. (1985) using the program Li93 (Li 1993). Relative rate tests were carried out according to the method of Li and Tanimura (1987). Statistical significance was tested with Student's *t* statistic (Zar 1984) and Bonferroni correction (Rice 1989; Holm 1979).

Southern and Northern Blot Analyses. Southern blot analysis of 10 μg of genomic DNA digested with *Eco*RI was performed as described in Moniz de Sá and Drouin (1996) with the last three 10-min posthybridization washes performed with 0.1 × SSC and 0.5% SDS at 42°C. Northern blot analysis of tomato pollen RNA (a kind gift from Sheila McCormick, Plant Genome Expression Center, Berkeley, CA) was performed under the same hybridization conditions. The Pot42 and Tom32 probes consisted of the complete cloned PCR fragments of these two genes. These two DNA fragments were used to probe two of the Southern blots previously used by Moniz de Sá and Drouin (1996). The Tac25 probe consisted of the 2.6-kb *Bam*HI fragment of the pTac25 plasmid and was used to probe a new Southern blot. This fragment contains the complete coding region of this gene and was gel isolated after digesting the pTac25 plasmid with both *Bam*HI and *Xba*I so as to obtain fragments that did not comigrate with the cloning vector (pUC, which is also 2.6 kb long). The pTac25 plasmid was a kind gift from David Lonsdale (John Innes Centre, Norwich, UK).

Results

Phylogenetic and Sequence Analyses

Both the neighbor-joining and parsimony trees of the first and second bases of codons of 52 angiosperm actin genes published in Moniz de Sá and Drouin (1996) showed that the Pot42, Pot65, Tom32, and Tac25 genes formed a cluster where the gene tree topology corresponded to the known species tree topology, i.e., ((potato, tomato) tobacco). These results therefore suggested that these genes were orthologous. Southern blot analyses show that these four genes are part of the same subfamily, but they are not strictly orthologous (see below).

Sequence analysis of the tomato actin gene Tom32 showed that it has lost its second intron while the potato actin genes Pot42 and Pot65 have both lost their second and third introns. In all cases the removal of the introns is precise, leaving the coding sequence in frame (results not shown). Tom32 is likely a pseudogene since it has two stop codons (at codons 65 and 225 of its actin coding region), and its first intron is only 63 bp long. A minimal intron size of 68 bp is thought to be required for efficient splicing of angiosperm transcripts (Goodall and Filipowicz 1990). Furthermore, the fact that our Tom32 probe did not hybridize to a Northern blot of tomato pollen RNA suggests that this gene is indeed not expressed in pollen (results not shown). The potato genes Pot42 and Pot65 are potentially functional but the proteins they code for are 373 and 378 amino acids long, rather than the 377-amino-acid-long proteins usually encoded by angiosperm actin genes. Pot42 contains a deletion of four amino acids (codons 233–236 of its actin coding region) and Pot65 contains an amino acid insertion (an extra lysine codon between codons 236 and 237 of its actin coding region). Whether or not the proteins encoded by these genes are functional is not known, but relative rate tests do show that the nonsynonymous sites of these two genes are not evolving at faster rates when compared to the tobacco Tac25 actin gene. The rates of nonsynonymous substitutions, and their standard errors, for the Pot42-Tac25 and Pot65-Tac25 comparisons are 1.52 ± 0.70 and 1.10 ± 0.74, respectively. These values are not significantly different from zero at the Bonferroni-corrected level of 2.5%. This suggests that these two genes might be functional. It should also be noted that intron loss does not necessarily lead to nonfunctional genes. The ACT2 actin gene of *A. thaliana,* which lost its first intron, has been shown to be highly expressed (see above).

Southern Blot Analyses

The Southern blot results shown in Fig. 1 indicate that the actin genes Pot42 and Tom32 are orthologous and that they are part of the same actin gene subfamily as Tac25. The Pot42 and Tom32 probes not only hybridize strongly to an *Eco*RI fragment in the digest of their respective species, but the most strongly hybridizing fragment they recognize in the other species also corresponds to that fragment. The Tac25 probe also hybridizes to the same potato and tomato fragments, but it is not the frag-

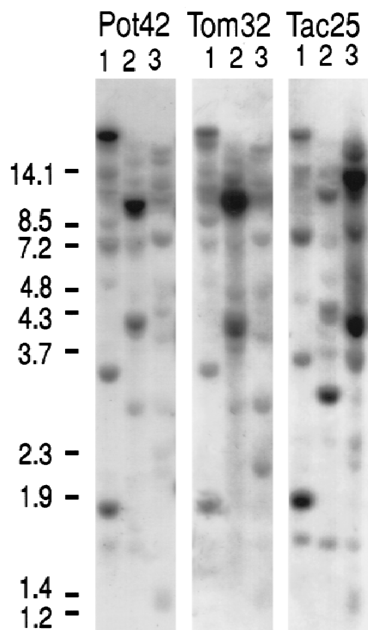

Fig. 1. Southern blot analysis of a group of related actin genes found in the potato, tomato, and tobacco genome. Each replicate blot contained 10 μg of *Eco*RI-digested genomic DNA from potato (*lanes 1*), tomato (*lanes 2*), and tobacco (*lanes 3*) and was probed with the DNA fragment indicated. Molecular size markers are indicated to the left of the autoradiograms (in kb). Note that the Southern blot used for the *Tac25* hybridization was slightly shorter than the other two.

ments to which it hybridizes most strongly. Whereas the Pot42 and Tom32 probes hybridize most strongly to a 30-kb and an 11-kb fragment, respectively, the Tac25 probe hybridizes most strongly to an 1.8-kb band in potato and a 3.2-kb band in tomato. On the other hand, the hybridization pattern of these three genes is different from that of the five other orthologous groups of genes previously identified between these three species (Moniz de Sá and Drouin 1996). These observations suggest that Pot42 and Tom32 are part of the same actin gene subfamily as the tobacco Tac25 actin gene. Although the potato and tomato gene are not strictly orthologous to the tobacco gene, these three genes likely represent closely related paralogous subfamily members. In fact, these hybridizations also show that within each species there is more than one sequence having relatively high similarity to the probes used. Since angiosperm actin genes have 1.1 kb of coding region and usually three introns that vary in size from 100 to 500 bp, they are typically at most 2.6 kb long (Moniz de Sá 1995). The hybridization of each probe to several restriction fragments larger than 2.6 kb therefore suggests that they recognize a small multigene subfamily within each species.

Discussion

We found three angiosperm actin sequences which did not possess the three introns typical of angiosperm actin genes. The tomato clone Tom32 lacked the second intron, whereas the two potato clones, Pot42 and Pot65, lacked both the second and third introns. The loss of the introns was precise such that the reading frame was not altered. This clearly represents a derived condition since over 60 other angiosperm actin gene sequences are now known, and all these genes have three introns in identical locations. Furthermore, the tobacco Tac25 actin gene is related to these genes and contains three introns (Thangavelu et al. 1993). These relationships, suggested by the phylogenies of Moniz de Sá and Drouin (1996), were confirmed by Southern blot analysis (Fig. 1).

The fact that both the Pot42 and Pot65 potato actin genes have lost their second and third introns is most parsimoniously explained by considering that they originated from the recent duplication of an ancestral potato actin gene, which had lost both introns. These two potato actin genes differ from one another by only 0.05 synonymous substitutions/site (results not shown). Given that the rate of synonymous substitutions of angiosperm genes has been calculated to be 7×10^{-9} substitutions/site/year, these two genes therefore diverged 3.5 MYA (Moniz de Sá and Drouin 1996). This date is more recent than the divergence time of potato and tomato, which has been estimated at 11.6 ± 3.6 MYA (Moniz de Sá and Drouin 1996).

There are therefore two different scenarios by which introns could have been lost in the potato and tomato genes. The second intron could have been lost before the divergence of tomato and potato, followed by the loss of the third intron in the potato lineage, or the loss of intron two in the tomato lineage was independent of the simultaneous loss of introns 2 and 3 in the potato lineage. It is not possible to determine which scenario is the most likely since both require two intron loss events.

Several cases have been described where intron loss is clearly a derived condition in the angiosperm nuclear genes, i.e., where variation in intron numbers is very unlikely to be the result of intron addition in closely related lineages. The evolution of legumin genes provides what is likely the most convincing example of repeated intron loss events in nuclear genes of angiosperms. In gymnosperms, these genes contain four introns. Not only was intron 4 lost early during angiosperm evolution, but the lineage leading to sunflower and oilseed rape subsequently lost intron 3, while the lineage leading to field bean and pea subsequently lost intron 1 (reviewed in Häger et al. 1996). Other examples where intron loss is clearly a derived condition include the loss of the second intron in the group 1 α-amylase genes of wheat and barley, as well as the loss of one intron in the kinase domain of the S-receptor kinase 3 gene in *Brassica oleracea* (Huang et al. 1990; Kumar and Trick 1993). Interestingly, legumin and group 1 α-amylase genes are expressed in seeds, whereas the S-receptor kinase genes are expressed in pistil, anther, stigma, leaf,

and root tissues (Shotwell and Larkins 1989; Huang et al. 1990; Stein et al. 1991; Glavin et al. 1994; Kumar and Trick 1994).

The most likely mechanism by which introns can be precisely deleted is by gene conversion with an intronless reverse-transcribed cDNA copy of one of their transcripts or that of a closely related gene. Such a mechanism was first proposed by Lewin (1983). Fink (1987) convincingly argued that this mechanism was responsible for the paucity of introns in *Saccharomyces cerevisiae,* and that the few remaining introns are almost always found at the 5′ end of the genes of this species. The same argument has recently been made to explain the paucity and asymmetrical location of introns in red algae genes (Liaud et al. 1995). The involvement of cDNAs in homologous gene conversion events resulting in the precise splicing out of yeast introns has been demonstrated experimentally (Derr et al. 1991; Melamed et al. 1992; Derr and Strathern 1993). Precise intron loss in several angiosperm mitochondrial genes has also been hypothesized to be due to gene conversion with a cDNA intermediate (Geiss et al. 1994, and references therein). In animals, the absence of an intron in the coding region of the rat preproinsulin I gene has been shown to be the result of the fact that this gene was generated by an RNA-mediated duplication-transposition event involving a transcript of the preproinsulin II gene (Soares et al. 1985).

Although reverse-transcriptase activity has not yet been demonstrated in angiosperms, both copia-like and gypsy-like retrotransposons have been found to be ubiquitous components of angiosperm genomes and could be the source of this enzyme (reviewed in Grandbastien 1992; Wessler et al. 1995). This enzyme could also be derived from plant retroviruses such as the cauliflower mosaic virus (Hull and Covey 1983). Although the precise origin of the reverse-transcriptase activity responsible for the formation of nuclear cDNA molecules is still not clear, the existence of an actin-processed pseudogene in the potato genome clearly shows that reverse transcription does occur in this species (Drouin and Dover 1987; Dornburg and Temin 1990; Luan et al. 1993).

The significance of our results lies in the facts that we observed two independent intron loss events among angiosperm actin genes which otherwise contain three introns and that these genes are related to a tobacco actin gene which has been shown to be expressed only in mature pollen (Tac25; Thangavelu et al. 1993). This is consistent with an intron loss mechanism which involves recombination with a cDNA derived from a transcript expressed in the germ line. It has been noted that if gene conversion goes through a cDNA intermediate, then genes that are highly expressed in the germ line will tend to be the main donors of genetic information (Drouin and Dover 1987; Melamed et al. 1992). In angiosperms, the germ line is derived from somatic tissue late in development. Therefore, mutations will be fixed if they occur in germ tissue, tissues leading up to it, or early in seed development. Our results also suggest that intron loss is more frequent in genes having high sequence similarity with genes expressed in the germ line. Since mammalian gene conversion studies have shown that the frequency of gene conversion increases with both the length and the similarity of the sequences involved (Rubnitz and Subramani 1984; Letsou and Liskay 1987; Liskay et al. 1987), one would expect that intron loss by gene conversion with cDNA molecules would occur more frequently with sequences having high similarity with these cDNA molecules.

The intron loss events reported here, and the other examples of intron loss in angiosperm nuclear genes described above, involved genes expressed in reproductive tissues or in seeds. This suggests that gene conversion with reverse-transcribed cDNAs that are produced in the germ line, tissues leading to it, or early in seed development can lead to intron loss in angiosperm nuclear genomes. However, we should note that there is now convincing evidence that intron loss is not sufficient to explain the observed distribution of introns in genes. In fact, recent evidence regarding the position of introns in genes of related species is difficult to explain by intron loss only. Convincing cases for intron insertions have been made for the genes coding for actin, tubulin, and triosephosphate isomerase, among others (Dibb and Newman 1989; Weber and Kabsch 1994; Logsdon et al. 1995; Kwiatowski et al. 1995; Bhattacharya and Weber 1997).

Acknowledgments. We thank David Lonsdale (John Innes Centre, Norwich, UK) for the pTac25 clone, Sheila McCormick (Plant Genome Expression Center, Berkeley, CA) for the Northern blot of tomato pollen RNA, as well as R.B. Meagher and an anonymous referee for their constructive comments on a previous version of this manuscript. This work was supported by a research grant from the Natural Sciences and Engineering Research Council of Canada to G. D.

References

An Y-Q, McDowell JM, Huang S, Mckinney EC, Chambliss S, Meagher RB (1996) Strong, constitutive expression of the *Arabidopsis* ACT2/ACT8 actin subclass in vegetative tissues. Plant J 10:107–121

Bhattacharya D, Weber K (1997) The actin gene of the glaucocystophyte *Cyanophora paradoxa:* analysis of the coding region and introns, and an actin phylogeny of eukaryotes. Curr Genet 31:439–446

Carroll RL (1988) Vertebrate paleontology and evolution. WH Freeman, New York

Carroll SL, Bergsma DJ, Schwartz RJ (1986) Structure and complete nucleotide sequence of the chicken α-smooth muscle (aortic) actin gene. J Biol Chem 261:8965–8976

Derr LK, Strathern JN (1993) A role for reverse transcripts in gene conversion. Nature 361:170–173

Derr LK, Strathern JN, Garfinkel DJ (1991) RNA-mediated recombination in *S. cerevisiae.* Cell 67:355–364

Dibb NJ, Newman AJ (1989) Evidence that introns arose at protosplice sites. EMBO J 8:2015–2021

Dornburg R, Temin HM (1990) cDNA genes formed after infection with retroviral vector particles lack the hallmarks of natural processed pseudogenes. Mol Cell Biol 10:68–74

Drouin G, Dover GA (1987) A plant processed pseudogene. Nature 328:557–558

Fink GR (1987) Pseudogenes in yeast? Cell 49:5–6

Geiss KT, Abbas GM, Makaroff CA (1994) Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. Mol Gen Genet 243:97–105

Glavin TL, Goring DR, Schafer U, Rothstein SJ (1994) Features of the extracellular domain of the S-locus receptor kinase from *Brassica.* Mol Gen Genet 244:630–637

Goodall GJ, Filipowicz W (1990) The minimum functional length of pre-mRNA introns in monocots and dicots. Plant Mol Biol 14:727–733

Grandbastien M-A (1992) Retroelements in higher plants. Trends Genet 8:103–108

Häger K-P, Müller B, Wind C, Erbach S, Fischer H (1996) Evolution of legumin genes: loss of an ancestral intron at the beginning of angiosperm diversification. FEBS Lett 387:94–98

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Huang N, Sutliff TD, Litts JC, Rodriguez RL (1990) Classification and characterization of the rice α-amylase multigene family. Plant Mol Biol 14:655–668

Hull R, Covey SN (1983) Replication of cauliflower mosaic virus DNA. Sci Prog 68:403–422

Kumar V, Trick M (1993) Sequence complexity of the S receptor kinase gene family in *Brassica.* Mol Gen Genet 241:440–446

Kumar V, Trick M (1994) Expression of the S-locus receptor kinase multigene family in *Brassica oleracea.* Plant J 6:807–813

Kusakabe T, Araki I, Satoh N, Jeffery WR (1997) Evolution of chordate actin genes: evidence from genomic organization and amino acid sequences. J Mol Evol 44:289–298

Kwiatowski J, Krawczyk M, Kornacki, Bailey K, Ayala FJ (1995) Evidence against the exon theory of genes derived from the triosephosphate isomerase gene. Proc Natl Acad Sci USA 92:8503–8506

Letsou A, Liskay RM (1987) Effect of the molecular nature of mutation on the efficiency of intrachromosomal gene conversion in mouse cells. Genetics 117:769

Lewin R (1983) How mammalian RNA returns to its genome. Science 219:1052–1054

Li W-H (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36:96–99

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Liaud M-F, Brandt U, Cerff R (1995) The marine red algae *Chondrus crispus* has a highly divergent β-tubulin gene with a characteristic 5′ intron: functional and evolutionary implications. Plant Mol Biol 28:313–325

Liskay RM, Letsou A, Stachelek JL (1987) Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115:161–167

Liu ZL, Zhu ZY, Roberg K, Faras A, Guise K, Kapuscinski AR, Hackett PB (1990) Isolation and characterization of the beta-actin gene of carp (*Cyprinus carpio*). DNA Seq 1:125–136

Logsdon JM, Tyshenko MG, Dixon C, Jafari JD, Walker VK, Palmer JD (1995) Seven newly discovered intron positions in the triosephosphate isomerase gene: evidence for the intron-late theory. Proc Natl Acad Sci USA 92:8507–8511

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605

McDowell JM, Huang S, McKinney EC, An Y-Q, Meagher RB (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana.* Genetics 142:587–602

Melamed C, Nevo Y, Kupiec M (1992) Involvement of cDNA in homologous recombination between Ty elements in *Saccharomyces cerevisiae.* Mol Cell Biol 12:1613–1620

Moniz de Sá M (1995) The evolution of angiosperm actin genes. PhD thesis, University of Ottawa, Ottawa, Ontario, Canada

Moniz de Sá, Drouin G (1998) Phylogeny and substitution rates of one form and six gymnosporin active genes (submitted)

Moniz de Sá M, Drouin G (1996) Phylogeny and substitution rates of angiosperm actin genes. Mol Biol Evol 13:1198–1212

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225

Rubnitz J, Subramani S (1984) The minimum amount of homology required for homologous recombination in mammalian cells. Mol Cell Biol 4:2253–2258

Sheterline P, Sparrow JC (1994) Actin. Protein Profile 1:1–121

Shotwell MA, Larkins BA (1989) The biochemistry and molecular biology of seed storage proteins. In: Marcus A (ed) The biochemistry of plant, vol 15. Academic Press, New York, pp 297–345

Soares MB, Schon E, Henderson A, Karathanasis SK, Cate R, Zeitlin S, Chirgwin J, Efstratiadis A (1985) RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. Mol Cell Biol 5:2090–2103

Stein JC, Howlett B, Boyes DC, Nasrallah ME, Nasrallah JB (1991) Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea.* Proc Natl Acad Sci USA 88:8816–8820

Stewart WN (1983) Paleobotany and the evolution of land plants. Cambridge University Press, Cambridge

Thangavelu M, Belostotsky D, Bevan MW, Flavell RB, Rogers HJ, Lonsdale DM (1993) Partial characterization of the *Nicotiana tabacum* actin gene family: evidence for pollen-specific expression of one of the gene family members. Mol Gen Genet 240:290–295

Weber K, Kabsch W (1994) Intron positions in actin genes seem unrelated to the secondary structure of the protein. EMBO J 13:1280–1286

Wessler SR, Bureau TE, White SE (1995) LTR-retrotransposons and MITEs: important players in the evolution of plant genomes. Curr Opin Genet Dev 5:814–821

Zar JH (1984) Biostatistical analysis. 2nd ed. Prentice-Hall