Tempo and Mode of Spliceosomal Intron Evolution in Actin of Foraminifera

Flakowski, Jérôme; Bolivar, Ignacio; Fahrni, José; Pawlowski, Jan Wojciech

How to cite

FLAKOWSKI, Jérôme et al. Tempo and Mode of Spliceosomal Intron Evolution in Actin of Foraminifera. In: Journal of Molecular Evolution, 2006, vol. 63, n° 1, p. 30–41. doi: 10.1007/s00239-005-0061-z

This publication URL: https://archive-ouverte.unige.ch/unige:118387

Publication DOI: 10.1007/s00239-005-0061-z

J Mol Evol (2006) 63:30–41
DOI: 10.1007/s00239-005-0061-z

JOURNAL OF MOLECULAR EVOLUTION

Tempo and Mode of Spliceosomal Intron Evolution in Actin of Foraminifera

Jérôme Flakowski, Ignacio Bolivar, José Fahrni, Jan Pawlowski

Department of Zoology and Animal Biology, University of Geneva, Sciences III, 30 Quai Ernest-Ansermet CH-1211 Geneva 4, Switzerland

Received: 17 March 2005 / Accepted: 2 February 2006 [*Reviewing Editor:* Dr. Debashish Bhattacharya]

Abstract. Spliceosomal introns are present in almost all eukaryotic genes, yet little is known about their origin and turnover in the majority of eukaryotic phyla. There is no agreement whether most introns are ancestral and have been lost in some lineage or have been gained recently. We addressed this question by analyzing the spatial and temporal distribution of introns in actins of foraminifera, a group of testate protists whose exceptionally rich fossil record permits the calibration of molecular phylogenies to date intron origins. We identified 24 introns dispersed along the sequence of two foraminiferan actin paralogues and actin deviating proteins, an unconventional type of fast-evolving actin found in some foraminifera. Comparison of intron positions indicates that 20 of 24 introns are specific to foraminifera. Four introns shared between foraminifera and other eukaryotes were interpreted as parallel gains because they have been found only in single species belonging to phylogenetically distinctive lineages. Moreover, additional recent intron gain due to the transfer between the actin paralogues was observed in two cultured species. Based on a relaxed molecular clock timescale, we conclude that intron gains in actin took place throughout the evolution of foraminifera, with the oldest introns inserted between 550 and 500 million years ago and the youngest ones acquired less than 100 million years ago.

Key words: Foraminifera — Phylogeny — Actin paralogues — Spliceosomal introns — Bayesian relaxed molecular clock

Introduction

The majority of nuclear protein-coding genes are interrupted by spliceosomal introns. They probably originated in the eukaryote stem lineage prior to the diversification of the earliest protists as suggested by their existence in all eukaryotic phyla (Fast and Doolittle 1999; Nixon et al. 2002). Since the discovery of introns, two dominant hypotheses have been formulated to explain their origin: (i) the intron-early hypothesis, linked to the exon-shuffling model, which assumed the existence of formative introns subsequently lost in prokaryotes (Doolittle 1978; Gilbert et al. 1986, 1997), and (ii) the intron-late hypothesis, associated with the insertional model of intron, which postulates their acquisition in preformed genes (Cavalier-Smith 1991; Palmer and Logsdon 1991; Cho and Doolittle 1997). The current tendency is to adopt a mixed viewpoint, which reconciles the two hypotheses, accepting that some introns are highly conserved while the others are species-specific (Wang et al. 2005). However, there is no agreement on the proportion of ancient versus recent introns (Fedorov et al. 2002; Qiu et al. 2004).

Recent analyses of complete eukaryotic genomes revealed that as many as 25–30% of intron positions are shared between eukaryotic kingdoms, suggesting a remarkable degree of intron conservation (Rogozin et al. 2003). Maximum likelihood analysis of the same data set showed that the rates of intron loss are an order

Abbreviations: ACT1, actin type 1; ACT2, actin type 2; ADP, actin deviating protein; ARP, actin-related protein; BS, bootstrap support; ML, maximum likelihood; Mya, million years ago; NAP, novel actin-like protein; SSU, small subunit rRNA; SD, standard-deviation of Bayesian relaxed molecular clock method.

Correspondence to: Jan Pawlowski; *email:* Jan.Pawlowski@zoo.unige.ch

of magnitude higher than the rates of intron gain (Roy and Gilbert 2005a, b). The prevalence of intron loss over gain was also demonstrated by multigene comparisons of intron positions in human, mouse, and rat (Roy et al. 2003), as well as in species of *Caenorhabditis* (Cho et al. 2004). On the other hand, the case studies of particular genes and lineages show a large number of recently inserted introns (Funke et al. 1999; Bhattacharya et al. 2000; Robertson et al. 2003). Some of these introns are well conserved within closely related species and were considered to be potentially valuable phylogenetic markers (Brady and Danforth 2004).

Here, we examined the proportion of ancient versus more recent introns in actin coding genes of foraminifera. This group of marine protists was chosen because their rich and well-preserved fossil record, spanning more than 550 million years, allows us to calibrate their phylogeny and to date the origin of introns. Foraminiferan actin was chosen because it has a clear phylogenetic signal and its gene possesses introns that are easily detectable. Actin is a highly conserved protein, ubiquitously expressed in eukaryotic cells, that is involved in diverse types of cell motility and cytoskeleton structure (Sheterline et al. 1999). Although actin frequently presents multiple paralogues resulting from duplication events (Vandekerckhove and Weber 1978; Carlini et al. 2000), its phylogenies agree most of the time with those based on the small subunit (SSU) rRNA and other genes (Baldauf et al. 2000). Actin serves as a particularly good phylogenetic marker of amoeboid protists, such as Amoebozoa (Fahrni et al. 2003), Cercozoa, and Foraminifera (Keeling 2001; Nikolaev et al. 2004). The actin gene family also comprises more divergent relatives like the actin-related proteins (ARPs) (Goodson and Hawse 2002) and novel actin-like proteins (NAPs) described in volvocalean algae (Kato-Minoura et al. 2003).

The results of our study confirm the presence of two actin paralogues in foraminifera and reveal the presence of a set of unconventional actin deviating proteins (ADPs), which are neither ARPs nor NAPs. We identified 24 intron positions, of which only 4 have been previously reported in other eukaryotes. We inferred phylogenetic distribution of introns in two actin paralogues and dated their origins based on a Bayesian relaxed molecular clock approach. Our analyses suggest that actin introns have been gained throughout the evolution of foraminifera and that most of them are not phylogenetically conserved.

Materials and Methods

DNA Extraction, PCR Amplification, Cloning, and Sequencing

Among the 27 species of foraminifera examined in this study, 2 (*Reticulomyxa filosa*, *Allogromia* sp. A) originated from laboratory cultures. The remaining species were isolated from environmental samples (Flakowski et al. 2005). The specimens were individually cleaned with a paintbrush and rinsed in several baths of sterile seawater prior to extraction. DNA was extracted by using either a guanidinium buffer (Chomczynski and Sacchi 1987) or a DNeasy Plant MiniKit (Qiagen). Each DNA extraction contained from 50 to 100 specimens of the same morphospecies. PCR amplifications, cloning, and sequencing were done as described elsewhere (Pawlowski et al. 1999). The sequences of PCR primers and their positions are given by Flakowski et al. (2005). We deposited the 105 new sequences reported in this paper in the GenBank/EMBL database under accession numbers (gb) AY763936–AY764025 and AY766188) and (gb) AY763390–AY763404, respectively.

Phylogenetic Analyses

The actin family tree was inferred using the maximum likelihood (ML) method (Felsenstein 1981) from an alignment comprising 75 sequences. Foraminiferan actin sequences, their eukaryotic homologues, the ARPs, and the NAPs were aligned using Clustal X (Thompson et al. 1994) and further revised by eye. To accommodate rate variations among sites, distances were computed under the WAG (Whelan and Goldman 2001) substitution model, assuming a gamma distribution with Γ_8 discrete categories + invariable sites. The ML analysis was performed using PhyML v2.4 (Guindon and Gascuel 2003). We let the program estimate the proportion of invariable sites and the shape of the gamma distribution. We kept the topology and optimized the branch lengths and rate parameters. The 100 nonparametric ML bootstraps were also calculated using PhyML.

The relations among foraminifera were inferred from an actin alignment comprising 9 actin type 1 (ACT1) and 25 actin type 2 (ACT2) sequences. *Gromia oviformis*, a sister group to foraminifera (Longet et al. 2004), was chosen as outgroup. The phylogenetic analysis was achieved using the same approach as described above, but we optimized the tree topology by forcing the monophyly of Allogromiidae, based on SSU phylogenies (Pawlowski et al. 2002). We compared the nonconstrained and the constrained topologies using TREE-PUZZLE 5.2 (Schmidt et al. 2002) to perform the Shimodaira–Hasegawa (1999) test, the two-sided Kishino–Hasegawa (1989) test, and the expected likelihood weight test (Strimmer and Rambaut 2002). All tests indicated that the two topologies were not significantly different. The optimized topology was used in the relaxed molecular clock analysis.

Distinction of Actin Deviating Sequences

To distinguish putative ADPs from conventional actin genes, we performed several analyses of their sequences: (1) comparing intron positions; (2) searching for insertions and deletions in the putative ADPs sequences; (3) evaluating evolutionary rates using a relative rate test with the RRTree 1.1 program (Robinson-Rechavi and Huchon 2000); (4) looking for the signature actin motifs PS00406, PS00432, and PS01132 on ScanProsite (www.expasy.org/tools/scanprosite/) (Gattiker et al. 2002); (5) detecting perturbed protein structures in the putative ADPs using secondary structure predictions with Network Protein Sequence Analysis (npsa-pbil.ibcp.fr/); and (6) analyzing tertiary structure predictions with GENO3D (geno3d-pbil.ibcp.fr/) (Combet et al. 2002) with the actin PDB template pdb2btfA-0 suggested by the program.

Divergence Time Estimation

We first tested the presence of a global molecular clock in the amino acid data sets with a chi-square test. The ln$L$ was obtained

from a nonconstrained method with ProML 3.6 and a constrained method with ProMLK 3.6 (Felsenstein 2004) under the JTT model with Γ_8 discrete categories + invariable sites estimated under TREE-PUZZLE 5.2. To localize the fast-evolving lineages, we applied a relative rate test to the two paralogues with the RRTree 1.1 program (Robinson-Rechavi and Huchon 2000).

For dating analysis, we used a Bayesian relaxed molecular clock method implemented in the MULTIDIVTIME package (Thorne et al. 1998; Kishino et al. 2001; Thorne and Kishino 2002). The ESTBRANCHES program estimated the branch lengths of the constraint topology, based on the hmm38C model (Goldman et al. 1998), in order to obtain the variance-covariance matrices for our data sets. Thereafter, we processed the obtained matrices with ESTBRANCHES applying the MULTIDIVTIME program. We chose to sample the Markov chains 10,000 times, with 100 cycles between each sample and burn-in after 100,000 cycles. The prior for the Brownian motion constant and the gamma distribution of the rate at the node were calculated from the median branch length according to the manual recommendations. *Gromia* was our ingroup root and we chose 820 million years ago (Mya; SD, 410) for the prior number of time units between the ingroup root and the present time, a date situated in the lower part of a time interval of 1150–690 Mya suggested in a previous study (Pawlowski et al. 2003). We used prior constraint on multiple calibration windows—lower and upper bound—dispersed across the tree to reduce potential regional effects in order to estimate the dates of the ingroup root and other nodes.

Intron Distribution

We refer to the intron using the convention [codon: phase] (Dibb and Newman 1989). To trace the distribution of actin introns through time, we used parsimony reconstruction methods to find ancestral positions (Schluter et al. 1997; Pagel 1999). We used the MESQUITE package 1.05 (Maddison and Maddison 2004), particularly the module allowing us to trace a character history that takes into account the branch length of the input timescale tree. We coded a matrix of intron presence (1) and absence (0) to conduct parsimony intron analysis.

We calculated the interval of time in which the introns were inserted based on the length of internal branches between the ingroup with the intron and the intronless sister group. The precision of the time interval depended on the standard deviation of the two internal nodes at each end of the insertion period. An ingroup where the species shared the same intron position defined a lower limit, whereas the sister group without the intron defined an upper limit. Coupling an upper and a lower limit was necessary to define a period where the intron was most probably inserted. In those cases where an intron was not present in all members of the ingroup, parsimony analysis was used to determine its ancestral position.

Exon-Intron Boundaries, Protosplice Site, and Intron Sequence Similarities

The exon-intron organization in foraminiferan actin genes was established by aligning existing cDNA sequences (gb: AJ132370–AJ132375) with the DNA sequences obtained in this paper. This procedure also permitted us to confirm the presence of the classical splice site GT-AG of U2-dependent spliceosomal introns. To obtain consensus sequences at the boundary of exons and introns, we used the Weblogo program (Crooks et al. 2004). The height of each stack of letters represented the information content (in bits) for a position, whereas the relative heights of letters within a stack reproduced their frequencies at that position. To identify the protosplice site consensus, we examined the sequence at exon-intron boundaries for all foraminifera and only then we looked at sequences whose outgroup did not posses the introns which are common to the ingroup.

We used the PRSS program, which is part of the FASTA package 3.3 (Pearson 2003), to estimate the sequence similarity between two introns. From the same package, we used the program LALIGN (Huang and Miller 1991) and LFASTA (Chao et al. 1992) from the infobiogen website (www.infobiogen.fr/services/menuserv.html) to align the introns. The first five and the last four nucleotides that contain the splice site were excluded from the analysis. Additionally, we checked the sequence similarity of introns by blasting them against the available database. We did not find any similarities between foraminiferan introns and known transposons.

Results

Characterization of Foraminiferan Actins

Phylogenetic analysis of 75 actin sequences, including 33 foraminifera, 36 other eukaryotes, and 6 NAP and ARP sequences, showed that all foraminiferan actins grouped together in a weakly supported clade (Fig. 1). The position of this clade, next to *Gromia oviformis* and the cercozoan *Chlorarachnion* sp., is consistent with preceding studies indicating the close relationships among Foraminifera, *Gromia*, and Cercozoa (Keeling 2001; Nikolaev et al. 2004). The tree was rooted at the ARP sequences and its general topology is congruent with the classical SSU-based eukaryotic phylogeny, with trichomonads, Euglenozoa, and Heterolobosea at the base of the tree (Baldauf et al. 2004). Some eukaryotic taxa, such as ciliates, known to possess extremely divergent actins (Philippe and Adoutte 1998), were not included. All main eukaryotic phyla were recovered, although with relatively weak bootstrap support.

The two major foraminiferan actin paralogues, ACT1 and ACT2, were supported by moderate bootstrap support (BS) values of 51% and 56%, respectively. Within these paralogues, we identified 16 sequences that are clearly distinct from the conventional foraminiferan actin types. We named these sequences *actin deviating proteins* (ADPs), based on their phylogenetic position and molecular characteristics. We initially associated the ADPs with ARPs (Goodson and Hawse 2002) or NAPs (Kato-Minoura et al. 2003). Yet, while the ADPs clearly branch within the foraminiferan clade, the ARPs + NAPs form a sister group to the conventional actins with 84% BS (Fig. 1). When we compared the similarity of the protein sequences within the same species, we found that the paralogues ACT1 and ACT2 differed by 11–14%, whereas the divergence between paralogues and the deviant actins reached up to 42%.

We identified seven characters that distinguish the ADPs from conventional actins (Table 1). All ADPs were characterized by an acceleration of evolutionary

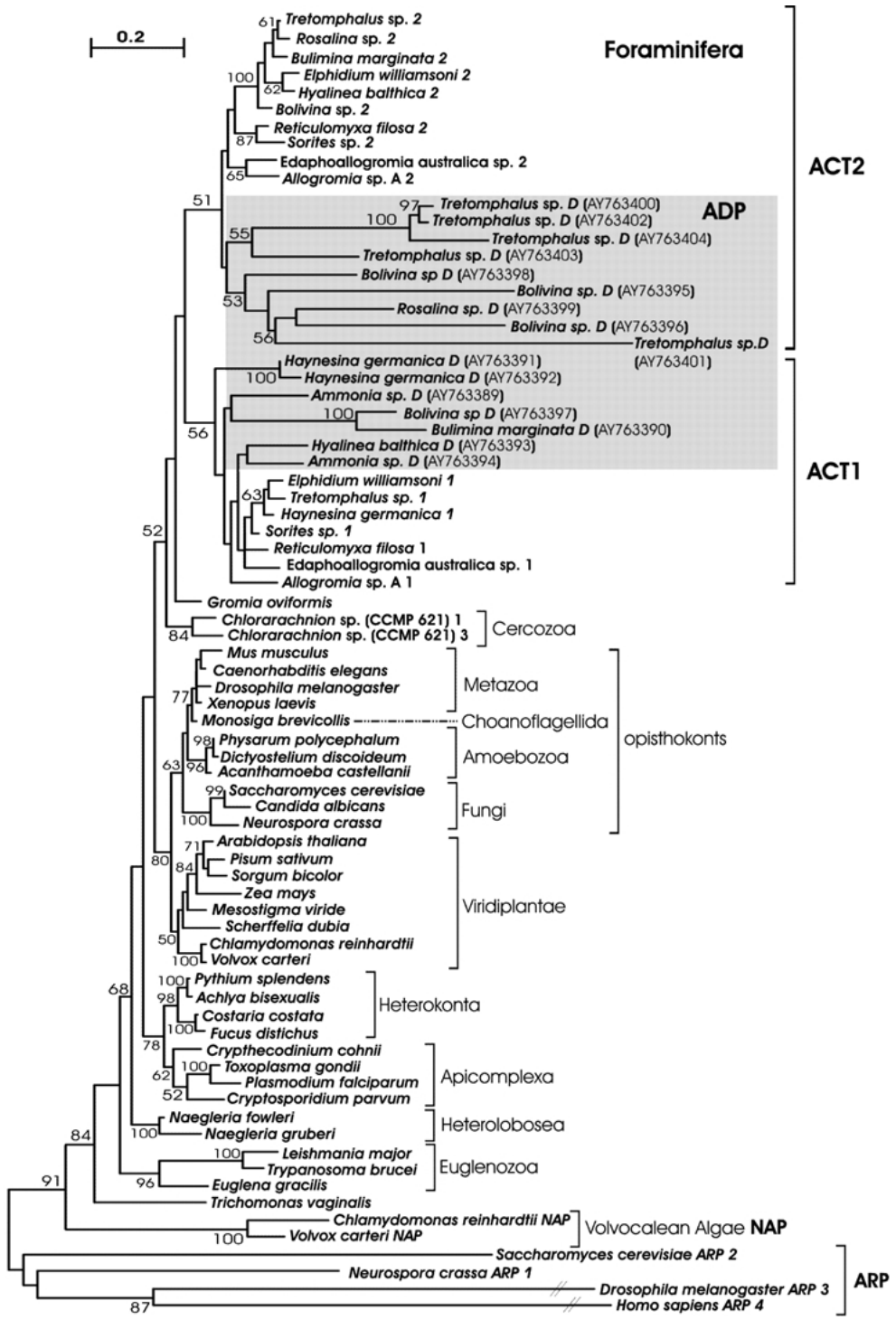

Fig. 1. Maximum likelihood actin family tree of eukaryotes rooted with the ARPs. The foraminiferan ADP sequences branch together with the two types of foraminiferan actin sequences (ACT1 and ACT2). The tree strongly supports the hypothesis that the foraminiferan ADPs belong to neither the ARPs nor the NAPs. The numbers on internal branches are bootstrap values of ML analyses, based on 100 replicates. Only values ≥50% are displayed.

rates ($P = 1e-7$), illustrated by their long branches (Fig. 1). The majority of ADPs showed indels: insertion up to 46 amino acids (sequences AY763395 and AY763398) and deletion up to 6 amino acids (for sequence AY763399). The actin motif research indicated that at least one of the three motifs was absent in the ADPs sequences. The analyses of the secondary structure prediction pattern indicated that the position and the length of the α-helix and ß-sheets were perturbed in the deviant sequence compared to the conventional sequence (supplementary material). The spatial organisation of tertiary structure prediction was mainly perturbed by the indels (supplementary material). Moreover, we found that only 4 of 13 DNA sequences of ADPs shared common introns with conventional actins (intron A and M) and that two introns were found only in ADPs ([52:1] and [216:0]).

Table 1. Characterization of the actin deviating proteins

Species	Sequence	Introns	Ins.	Del.	Rate	Motifs	2nd	3nd
Ammonia sp. (cDNA)	AY763389	—	—	—	x	x	x	—
Ammonia sp.	AY763394	M, ACT1[a]	—	x	x	x	x	x
Bolivina sp.	AY763395	—	x	—	x	x	x	x
Bolivina sp.	AY763396	[52:1]	x	—	x	x	x	x
Bolivina sp.	AY763397	—	—	x	x	x	x	x
Bolivina sp.	AY763398	A, ACT1[b]	x	—	x	x	x	x
Bulimina marginata	AY763390	—	—	x	x	x	x	x
Haynesina germanica	AY763391	M, ACT1[a]	x	—	x	x	x	x
Haynesina germanica	AY763392	M, ACT1[a]	x	—	x	x	x	x
Hyalinea balthica	AY763393	[216 :0]	x	—	x	x	x	x
Rosalina sp.	AY763399	—	x	x	x	x	x	x
Tretomphalus sp.	AY763400	—	x	—	x	x	x	x
Tretomphalus sp.	AY763401	—	x	x	x	x	x	x
Tretomphalus sp.	AY763402	—	x	—	x	x	x	x
Tretomphalus sp.	AY763403	—	—	x	x	x	x	x
Tretomphalus sp.	AY763404	—	x	x	x	x	x	x

Note. To distinguish the putative ADPs from conventional actin, we performed several analyses of their sequences and chose seven deviant characters: absence of common intron position between ADPs and conventional actins, insertions (Ins.), deletions (Del.), evolutionary rates, actin motifs, and perturbed secondary second and third structure predictions. (—) Absence of the deviant character; (x) presence of a deviant character.

[a] Intron in common with the paralogue of the same species.

[b] Intron in common with the sequence of Reophax sp. (gb: AY764009, AY764010).

Description of Foraminiferan Actin Introns

We identified 22 intron positions in foraminiferan conventional actins (Fig. 2), plus the 2 positions [52:1] (gb: AY763396) and [216:0] (gb: AY763393) in foraminiferan ADPs. Among these 24 positions, 20 were specific to foraminifera, while 4 were shared with the 64 distinct positions previously reported in other eukaryotes (Bhattacharya and Weber 1997; Qiu et al. 2004). The set of 64 positions did not comprise the 2 intron positions in *Gromia oviformis* (gb: AY571669) and the 2 intron positions in *Cercomonas* sp. actin sequences (gb: AF363534). The introns shared by foraminifera and other eukaryotes were the introns A and D present in *Caenorhabditis elegans* (gb: CAA34717), the intron T in *Cercomonas* sp. (gb: AF363534), and the intron U found in the ascidian *Halocynthia roretzi* (gb: BAA08112) and in volvocalean algae.

In general, the foraminiferan actin introns were small (mean, 89 nucleotides; range, 57–303 nucleotides) and dispersed along the whole sequence. Using Weblogo (supplementary material), we identified the consensus 5′–3′ splice-site sequence characteristic for all foraminifera (5′GTWW–WYAG3′), as well as the splice-site sequences specific for Rotaliida (5′GTATG–WYAG3′) and for the Allogromiidae (5′GTWW–WTAG3′). The phase distribution of extant intron (eight of each phase) was obviously not as biased as generally observed in eukaryotic introns (5:3:2) but was closer to the approximately uniform phase distribution of fungi than to other eukaryotes (Qiu et al. 2004). We then searched for the protosplice site MAG|R (Dibb and Newman 1989) in which intron insertion should take place. By using Weblogo (supplementary material), we examined exons at each end of the intron for all foramineran sequences and only then examined those sequences whose outgroup did not possess introns common to the ingroup. In the first case, the signal was weak (5′G reach 0.7 bits) and the consensus was WG|D. In the second case the signal was stronger (5′G reach 0.9 bits) and the obtained consensus MWG|G was nearer to the protosplice site sequence.

Phylogenetic Distribution of Introns

To provide the phylogenetic framework for the study of intron distribution, we analyzed 34 conventional foraminiferan actins, including 9 sequences of ACT1 and 25 sequences of ACT2 (Fig. 2). Given the lack of phylogenetic signal in the ADPs, we excluded them from our analyses. The mapping of introns on the phylogenetic tree of foraminifera allowed us to define their taxonomic range and specificity (Fig. 2). Among 22 introns identified in foraminiferan conventional actins, 3 showed a clear phylogenetic signal and were lineage-specific, with the intron F present in 10 of 12 sequenced Rotaliida ACT2, the intron M localized in 2 of 3 sequenced Rotaliida ACT1, and the intron C characteristic for ACT2 in Allogromiidae. The 9 other introns (A, D, H, I, L, P, S, T, and U) could also be lineage-specific, but our taxon sampling was insufficient to verify their presence in more than a single species. In Rotaliida and in Allogromiidae, the two lineages with relatively good taxonomic sam-

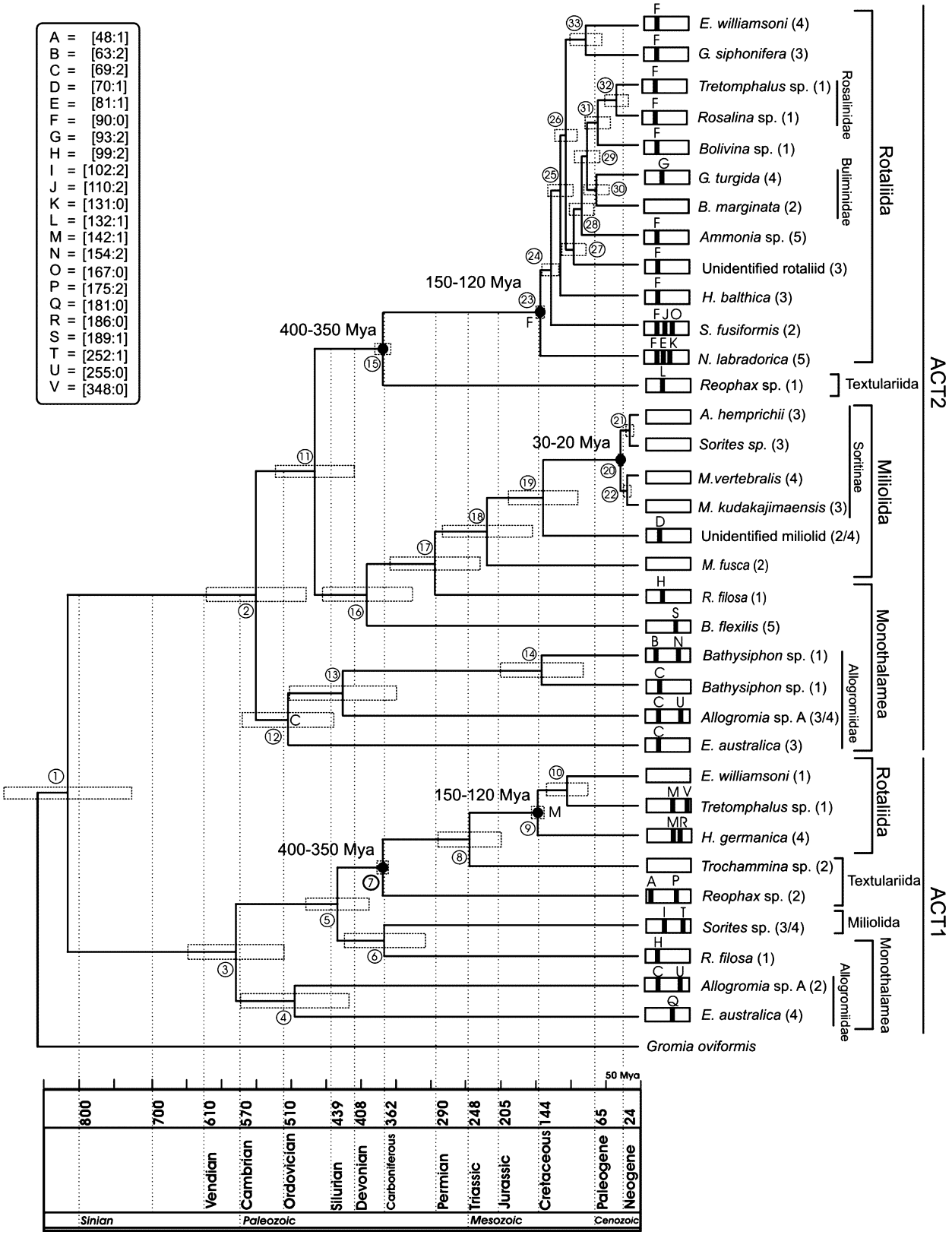

Fig. 2. Maximum likelihood chronogram of foraminiferal actin paralogues rooted with *Gromia oviformis*, the closest known protist to the foraminifera. The paralogue ACT1 comprises 9 species and the ACT2 comprises 25 species (7 species in common). The nodes with a spot indicate the calibration points. The intron ancestral states are indicated on the internal nodes with their corresponding letter. Intron positions are indicated on the rectangle on the left of the tree. The standard deviation is represented by a rectangle at each node. Small rectangles between the terminal nodes and the species names indicate the approximate position of the intron. The timescale is drawn at the bottom of the tree (one gradation = 50 Mya).

pling, we identified 10 species-specific introns (B, E, G, K, N, J, O, Q, R, and V). Some taxonomic groups, such as the Soritinae, for which four species were examined, seemed to lack introns in one of the paralogues (ACT2). Moreover, introns were not found in some other species (*E. williamsoni* and *Trochammina* sp. ACT1, *B. marginata* and *M. fusca* ACT2) as well as in some sequenced copies (*Allogromia* ACT2, unidentified miliolid ACT2, *Sorites* ACT1). The only clear cases of intron loss were observed for intron F in the ACT2 of Buliminidae (*B. marginata* and *G. turgida*), intron M in *E. williamsoni* ACT1, and intron C in one of the copies of *Bathysiphon* sp. ACT2. Interestingly, among seven species for which two paralogues were sequenced, two species (*Allogromia* sp. A and *R. filosa*) possessed the same introns (C + U and H, respectively) in both paralogues.

In order to test whether the introns share similar sequences, we compared the introns within and between foraminiferan species belonging to the same clades (Table 2). We did not find clear relationships between the phylogenetic proximity of species and the similarity of their introns. The intron F, which characterizes the paralogue ACT2 of Rotaliida, displayed a high sequence similarity only between *S. fusiformis* and an unidentified rotaliid, which are not directly related in our tree. In other Rotaliida, including the closely related *Rosalina* and *Tretomphalus*, the sequences of intron F were quite divergent. Among the highly similar introns (Table 2), we found introns localized at different positions (distance, 77 aa) in the same paralogue (*S. fusiformis* ACT2) and introns present in different positions in the paralogues of the same species (*Reophax* sp.—distance, 84 aa; *Ammonia* sp.—distance, 52 aa). High sequence similarities were also observed between introns present in the same position in the different paralogues of *Allogromia* sp. (introns C and U) and *R. filosa* (intron H), as well as between introns at the same position in conventional and deviant actins of *Ammonia* sp. (intron M).

Timescale of Intron Origins

The excellent fossil record of foraminifera allows us to calibrate the phylogenetic tree and to infer the divergence dates of the major lineages. We used the radiations of Textulariida (400−350 Mya), Rotaliida (150−120 Mya), and Soritinae (30−20 Mya) as calibration points according to well-recognized fossil dates (Haynes 1981; Loeblich and Tappan 1988; Ross and Ross 1991; Culver 1993). Because the presence of a global molecular clock in the amino acid data sets was rejected by the chi-square test ($p > 0.05$), we used a Bayesian relaxed molecular clock approach.

According to our data, the duplication of the actin gene in the ancestor of the foraminifera took place in the Neoproterozoic about 814 ± 89 Mya (Fig. 2).

Table 2. Intron sequence similarities of foraminiferan actins within species and between species belonging to the same clades

Species		ACT		Intron position		I. length		PRSS *p*-value	LALIGN	AL	LFASTA	AL
Allogromia sp. A (AY763944)	*Allogromia* sp. A (AY763940)	1	2	U[255:0]	U[255:0]	76	89	2.94127e-09	68.8 %	75	65.789%	76
S. fusiformis (AY764020)	*S. fusiformis* (AY764020)	2	2	F[90:0]	O[167:0]	59	48	6.0121e-05	73.2%	41	72.973 %	37
Reophax sp. (AY764008)	*Reophax* sp. (AY764009)	2	1	L[132:1]	A[48:1]	84	60	2.16893e-05	79.4 %	34	79.412%	34
R. filosa (AY763945)	*R. filosa* (AY763946)	1	2	H[99:2]	H[99:2]	60	82	0.000442289	67.2 %	61	66.102%	59
Allogromia sp. A (AY763944)	*Allogromia* sp. A (AY763941)	1	2	U[255:0]	U[255:0]	76	58	0.000135178	69.1%	55	69.091%	55
Unidentif. rotalid (AY763936)	*S. fusiformis* (AY764020)	2	2	F[90:0]	F[90:0]	92	59	0.00097921	61.5%	39	61.538%	39
Ammonia sp. (AY763951)	*Ammonia* sp. (AY763394)	1	D	M[142:1]	M[142:1]	77	79	7.96634e-05	78.8 %	33	80.769%	26
Ammonia sp. (AY763951)	*Ammonia* sp. (AY766188)	1	2	M[142:1]	F[90:0]	77	57	0.00326782	63.5 %	52	68.2%	22
Ammonia sp. (AY766188)	*Ammonia* sp. (AY763394)	2	D	F[90:0]	M[142:I]	75	79	0.00335232	66.7%	57	63.636%	55
Allogromia sp. A (AY763944)	*Allogromia* sp. A (AY763939)	1	2	C[69:2]	C[69:2]	78	95	0.00396142	66.2 %	77	71.053%	38
Rosalina sp. (AY764011)	*Tretomphalus* sp. (AY764006)	2	1	F[90:0]	M[142:I]	74	132	0.0206379	68.3 %	60	67.241 %	58

Note. Sequence similarities tested with the PRSS, LALIGN and LFASTA programs. We excluded the first five and the last four nucleotides which contain the splice signal (9 nucleotides must be added for the total length). Each intron sequence is described with its related species, GenBank accession number, actin type (1:ACT1, 2:ACT2, D: ADP), the intron position and the intron length (L Length). For PRSS, we indicate the P-value of a possible similarity by chance and for LALIGN and LFASTA the aligned sequence length (AL) in nucleotides and the percent of similarity between the aligned introns.

Table 3. Chronogram dates and ancestral intron positions

Node	Related event or clade	BV%	Date ± SD	Intron	
				Position	Parisomony
1	Duplicate Event	—	813 ± 89	—	—
2	Extant Foraminifera (2)	100	544 ± 73	C	a
3	Extant Foraminifera (1)	100	573 ± 72	—	—
4	Clade M (1)	48	489 ± 79	C	a
				Q, U	a
5	—	79	428 ± 47	—	—
6	—	9	362 ± 60	H, I, T	a
7	Textulariida + Rotaliida (1)*	57	363 ± 8	A, P	a
8	—	35	241 ± 46	M	a
9	Rotaliida (1)*	22	143 ± 11	M	a/p
				R	a
10	—	53	101 ± 30	V	a
11	—	70	460 ± 60	—	—
12	Clade M (2)	—	499 ± 74	C	p
13	—	31	421 ± 80	U	a
14	—	97	137 ± 67	B, N	a
15	Textulariida + Rotaliida (2)*	100	363 ± 11	F	a
				L	a
16	—	36	386 ± 66	S	a
17	—	50	289 ± 67	H	a
18	Miliolida + *Miliammina* (2)	51	215 ± 62	—	—
19	Miliolida (2)	57	136 ± 50	D	a
20	Soritinae (2)	99	26 ± 3	—	—
21	—	90	16 ± 7	—	—
22	—	39	13 ± 7	—	—
23	Rotaliida (2)*	43	140 ± 5	F	p
				E, K	a
24	—	28	125 ± 13	J, O	a
25	—	12	111 ± 16	F	p
26	—	2	103 ± 16	”	”
27	—	2	92 ± 17	”	”
28	—	7	82 ± 17	”	”
29	—	12	72 ± 17	”	”
30	Buliminidae (2)	15	61 ± 17	F, G	a
31	—	12	57 ± 17	F	p
32	Rosalinidae (2)	75	31 ± 15	”	”
33	—	58	75 ± 21	”	”

Note. (1) ACT1; (2) ACT2. BV, bootstrap value; SD, standard deviation; Position, intron position (A to V); Parsimony, parsimony model; a, intron absence; p, intron presence.
*Calibration point.

The diversification of extant foraminifera occurred during the Early Cambrian, about 572 ± 71 Mya according to analyses of the ACT1 data and about 545 ± 73 Mya using ACT2. The family Allogromiidae, representing the organic walled single-chamber foraminifera that are considered to be the most ‘‘primitive’’ forms, was dated with the ACT1 at 488 ± 79 Mya and with the ACT2 at 500 ± 74 Mya, which corresponds to the end of the Cambrian. For other lineages, there was relatively good congruence between dates inferred from the molecular chronogram and those inferred from the fossil record.

The patterns of the ancestral intron positions were obtained by parsimony analyses (Table 3), which allowed us to estimate the upper limit for the appearance time of all introns and the lower limit for three of them (F, M, and C2) (Fig. 3). The upper limit values ranged from 61 ± 17 Mya (intron G) to 544 Mya (intron C2). The majority of introns found in Rotaliida (except introns F and M), two introns in *Bathysiphon* sp. (B and N), and one intron in an unidentified miliolid (D) had upper limits that did not exceed ~150 Mya (Fig. 3). For nine introns (A, F, H1, H2, I, L, P, S, T) the upper limit ranged between 300 and 400 Mya; for intron U2 it was above 400 Mya, and three introns (C1, Q, U1) had an upper limit of about 500 Mya. The large number of introns with almost-identical upper limits may reflect a lack of sister-group sequences due to insufficient taxon sampling.

To reduce the interval between upper and lower limits, closely related species were necessary, as shown by the example of the Rotaliida-specific introns M and F present in ACT1 and ACT2, respec-

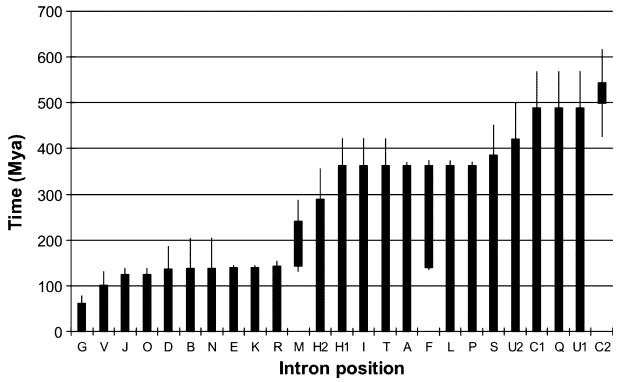

Fig. 3. Plot showing time intervals during which introns were probably inserted. In the case of introns common to the two paralogues, a number indicates the paralogue holding the intron. The time intervals are represented by black boxes and the thin lines correspond to their standard deviation. For three introns (M, F, C2) the lower and upper limits are indicated, whereas for the other introns only the upper limits were established. The introns are arranged according to their time of origin and not to their position along the foraminiferan actin paralogues.

tively. For intron F, the upper limit was estimated as 363 ± 11 Mya, which corresponds to the divergence time of *Reophax* sp. and Rotaliida in the fossil record. On the other hand, intron M, absent in the *Trochammina* sequence that branches as the sister group to Rotaliida (143 ± 11 Mya) in the ACT1 phylogeny, had an upper limit estimated to be 241 ± 46 Mya. The period of insertion was reduced nearly twofold (from 220 to 122 Mya) due to the presence of a closer sister group to Rotaliida in the ACT1.

Discussion

Origin of ADPs

The ADPs do not seem to be restricted to foraminifera. An analysis of cercozoan actin sequences suggested that at least some of them, such as the two *Cercomonas* actins (act2—gb, AF363537; and act3—gb, AF363538) can be considered ADPs. The ADPs are certainly not a third conventional paralogue since they do not form a monophyletic group with an apparent taxonomic signal. They are neither ARPs nor NAPs, because they are more similar to other foraminiferan actins than to ARPs (30–40%) and because the NAPs do not display indels (Kato-Minoura et al. 2003). Finally, they cannot be considered pseudogenes, because their sequences lack the stop codons.

Probably, the ADPs are produced by a dynamic process based on duplication of conventional actins followed by a rapid postduplication divergence of one of the copies (Conant and Wagner 2003). The high rate of actin duplication in foraminifera is supported by the existence of numerous closely related actin sequences (up to five copies of one paralogue; the number of copies is indicated in Fig. 2) for each species. Therefore, the duplication of one of the paralogues could possibly produce a conventional copy and an ADP after a given lapse of time. A good illustration of this phenomenon is provided by four ADPs of *Tretomphalus* sp., which show three ranges of divergences (Fig. 1). This hypothesis is also supported by the fact that some ADPs carry introns at the same position as the classical paralogues; for example, *H._germanica* (gb: AY763391 and AY763392) and *Ammonia* sp. (gb: AY763394) have the intron M [142:1], which is also held by their conventional actins and which defines the Rotaliida clade in the ACT1.

Actin Intron Evolution

The majority (83%) of the actin introns identified in this study are specific to foraminifera, suggesting that they have been gained after the divergence of the group from *Gromia*-like lineage (Longet et al. 2004). According to our chronogram, this divergence occurred more than 800 Mya. However, the large gap of more than 150 Mya between the duplication event (800 Mya) and the diversification of the two paralogues (645–501 Mya for ACT1 and 617–471 Mya for ACT2) could be the result of a postduplication acceleration effect (Jordan et al. 2004). Most probably, the foraminifera diverged from other eukaryotes about 690 Mya, as suggested by a revised dating of their SSU rDNA phylogeny (unpublished data). According to our timescale, the introns were gained throughout the evolution of foraminifera, with the oldest intron inserted about 550 Mya and the youngest inserted less than 65 Mya. Some introns might even be much younger, because the upper limits of their insertion dates are probably overestimated due to the lack of sequence data for closely related taxa.

The relatively recent origin of foraminiferan introns contradicts the "intron-early" hypothesis and its arguments for the large proportion of ancestral introns and their conservation (Roy et al. 2003; Roy and Gilbert 2005a, b). According to these studies, more than 25% of introns are shared between different eukaryotic kingdoms. In contrast, we found only four introns that are shared between foraminifera and other eukaryotes. Among them, only one (intron U) was found in more than one eukaryotic species. Because none of these four introns is phylogenetically conserved in foraminifera, it seems unlikely that they have an ancestral origin, for that would imply their loss in the majority of eukaryotes and foraminifera. Given the unique character and the phylogenetic distinctiveness of these introns, it is much more parsimonious to interpret them as result of parallel gains, as in the case of an intron identified in the xanthine dehydrogenase gene of plants and animals (Tarrio et al. 2003).

In general, intron gain ($n = 24$) by far outnumbers intron loss ($n = 3$) in foraminiferan actin. This finding is in striking opposition to massive intron losses suggested by interkingdom genome analyses (Rogozin et al. 2003; Roy and Gilbert 2005) and comparison of closely related species of nematodes (Cho et al. 2004). In the later study, the frequent loss of introns was explained by their small size. However, the foraminiferan data reported here fail to reveal any significant differences in size between the three introns that have been lost and the others.

An interesting case of recent intron gain is the presence of identical introns in both paralogues of *Allogromia* sp. (introns C and U) and *R. filosa* (intron H). The high similarity of their sequences suggests that the three introns have been transferred from one paralogue to the other relatively recently, probably as a result of gene conversion. Examples of such gene conversion have been reported previously for actin (Moniz de Sa and Drouin 1996) and other genes (McGuigan et al. 2004). Since these two species derive from long-established laboratory cultures, the process of genetic homogenization and gene conversion could be facilitated by the absence of environmental constraints and continuous clonal reproduction. Intron transfer within and between paralogues is also suggested by intron sequence similarity in some other species (Table 2). The exact mechanism of these transfers remains unclear but it is unlikely that they originated by gene conversion because of the significant distance between transferred introns.

The high number of intron gains in foraminiferan actin is surprising given the conserved character of this molecule and limited number of intron gain and loss in other eukaryotic phyla (Bagavathi and Malathi 1996). It is possible that this increase in intron insertions is related to the duplication of the actin gene during the early stage of foraminiferan evolution. It is well recognized that gene duplication can act as a factor accelerating evolutionary rates. For example, accelerated rates of intron gain and loss have been observed in duplicate genes of *Plasmodium* (Castillo-Davis et al. 2004). The increased number of intron insertions due to gene duplication has also been shown by genomic analysis of paralogous gene families (Babenko et al. 2004). Alternatively, the observed increase in intron insertion in foraminiferan actin genes could be related to the presence of multiple protosplice sites (Sadusky et al. 2004; Sverdlove et al. 2004). Our data confirm the presence of such protosplice sites, although their sequence pattern slightly differs from the classical consensus MAG(G (Dibb and Newman 1989). Such heterogeneity was also observed among fungi, animals, and plants (Qiu et al. 2004).

It has been suggested that massive intron gain and loss occur during transitional periods of evolutionary history (Babenko et al. 2004). This could also be true for foraminifera, as illustrated by similar upper limits of intron origins (Fig. 3) corresponding to major transitions in foraminiferan evolution, such as the radiation of monothalamous lineages (~500 Mya), the emergence of the multichamber lineages (~350 Mya), and the radiation of rotaliids (~150 Mya). However, much larger taxon sampling would be necessary to ensure that this is not an artifact due to a lack of sister groups for many taxa.

To conclude, our study shows that the majority of introns have been gained after the duplication of foraminiferan actin. Compared to other eukaryotic actins, intron turnover in foraminifera is very high, with intron gain by far outnumbering intron loss. Few introns are phylogenetically conserved, but most of them seem to correspond to recent insertions in single lineages. It remains to be demonstrated whether or not this massive intron gain is a result of relaxed selection, which could also explain the origin of numerous ADPs found in foraminifera or, instead, results from positive selection related to a particular function of foraminiferan actin.

Acknowledgments. We would like to thank Elisabeth Alve, Sam Bowser, Colomban de Vargas, Evelyne Goubert, Maria Holzmann, Valérie Le Cadre, David Longet, Xavier Pochon, Stephane Polet, Magali Schweizer, and Tom Wilding for their help in collecting the foraminifera. We are grateful to Juan Montoya and Sam Bowser for providing constructive remarks on the manuscript. This research was supported by Swiss National Science Foundation Grant 3100A0-100415.

References

Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV (2004) Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res 32:3724–3733

Bagavathi S, Malathi R (1996) Introns and protein revolution—an analysis of the exon/intron organisation of actin genes. FEBS Lett 392:63–65

Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. Science 290:972–977

Baldauf SL, Bhattacharya D, Cockrill J, Hugenholtz P, Pawlowski J, Simpson AGB (2004) The tree of life, an overview. In: Cracraft J, Donoghue MJ (eds) Assembling the tree of life. Oxford University Press, Oxford, pp 43–75

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

Bhattacharya D, Lutzoni F, Reb V, Simon D, Nason J, Fernandez F (2000) Widespread occurrence of spliceosomal introns in the rDNA genes of Ascomycetes. Mol Biol Evol 17:1971–1984

Brady SG, Danforth BN (2004) Recent intron gain in elongation factor-1a of colletid bees (Hymenoptera: Colletidae). Mol Biol Evol 21:691–696

Carlini DB, Reece KS, Graves JE (2000) Actin gene family evolution and the phylogeny of coleoid cephalopods (Mollusca: Cephalopoda). Mol Biol Evol 17:1353–1370

Castillo-Davis CI, Bedford TBC, Hartl DL (2004) Accelerated rates on itron gain/loss and protein evolution in duplicate genes in human and mouse malaria parasites. Mol Biol Evol 21:1422–1427

Cavalier-Smith T (1991) Intron phylogeny: a new hypothesis. Trends Genet 7:145–148

Chao KM, Pearson WR, Miller W (1992) Aligning two sequences within a specified diagonal band. Comput Appl Biosci 8:481–487

Cho G, Doolittle RF (1997) Intron distribution in ancient paralogues supports random insertion and not random loss. J Mol Evol 44:573–584

Cho S, Jin S-W, Cohen A, Ellis RE (2004) A phylogeny of Caenorhabditis reveals frequent loss of introns during nematode evolution. Genome Res 14:1207–1220

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

Combet C, Jambon M, Deleage G, Geourjon C (2002) Geno3D: automatic comparative molecular modelling of protein. Bioinformatics 18:213–214

Conant GC, Wagner A (2003) Asymmetric sequence divergence of duplicate genes. Genome Res 13:2052–2058

Crooks GE, Hon G, Chandonia J-M, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

Culver SJ (1993) Foraminifera. In: Lipps J (ed) Fossil prokaryotes and protists. Blackwell Scientific, Boston, pp 203–247

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

Doolittle WF (1978) Genes in pieces: Were they ever together? Nature 272:581

Fahrni JF, Bolivar I, Berney C, Nassonova E, Smirnov A, Pawlowski J (2003) Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. Mol Biol Evol 20:1881–1886

Fast NM, Doolittle WF (1999) *Trichomonas vaginalis* possesses a gene encoding the essential spliceosomal component, PRP8. Mol Biochem Parasitol 99:275

Fedorov A, Merican AF, Gilbert W (2002) Large-scale comparison of intron positions among animal, plant, and fungal genes. Proc Natl Acad Sci USA 99:16128–16133

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (2004) PHYLIP (Phylogeny Inference Package), Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Flakowski J, Bolivar I, Fahrni J, Pawlowski J (2005) Actin phylogeny of foraminifera. J Foraminiferal Res 35:93–102

Funke RP, Kovar JL, Logsdon JM, Corrette-Bennett JC, Strauss DR, Weeks DP (1999) Nucleus-encoded, plastid-targeted acetolactate synthase genes in two closely related chlorophytes, Chlamydomonas reinhardtii and Volvox carteri: phylogenetic origins and recent insertion of introns. Mol Gen Genet 262:12–21

Gattiker A, Gasteiger E, Bairoch A (2002) ScanProsite: a reference implementation of a PROSITE scanning tool. Appl Bioinform 1:107–108

Gilbert W, de Souza SJ, Long M (1997) Origin of Genes. Proc Natl Acad Sci USA 94:7698–7703

Gilbert W, Marchionni M, McKnight G (1986) On the antiquity of introns. Cell 46:151–153

Goldman N, Thorne JL, Jones DT (1998) Assessing the impact of secondary structure and solvent accessibility on protein evolution. Genetics 149:445–458

Goodson HV, Hawse WF (2002) Molecular evolution of the actin family. J Cell Sci 115:2619–2622

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Haynes JR (1981) Foraminifera. Macmillan, London

Huang X, Miller W (1991) A time-efficient, linear-space local similarity algorithm. Adv Appl Math 12:337–357

Jordan IK, Wolf YI, Koonin EV (2004) Duplicated genes evolve slower than singletons despite the initial rate increase. BMC Evol Biol 4:22

Kato-Minoura T, Okumura M, Hirono M, Kamiya R (2003) A novel family of unconventional actins in volvocalean algae. J Mol Evol 57:555–561

Keeling PJ (2001) Foraminifera and Cercozoa are related in actin phylogeny: Two orphans find a home? Mol Biol Evol 18:1551–1557

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29:170–179

Kishino H, Thorne JL, Bruno WJ (2001) Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol Biol Evol 18:352–361

Loeblich AR, Tappan H (1988) Foraminiferan genera and their classification. Van Nostrand Reinhold, New York

Longet D, Burki F, Flakowski J, Berney C, Polet S, Fahrni J, Pawlowski J (2004) Multigene evidence for close evolutionary relations between Gromia and Foraminifera. Acta Protozool 43:303–311

Maddison WP, Maddison DR (2004) Mesquite: a modular system for evolutionary analysis. Software http://mesquiteproject.org.mesquite

McGuigan K, Phillips PC, Postlethwait JH (2004) Evolutiion of sarcomeric myosin heavy chain genes: evidence from fish. Mol Biol Evol 21:1042–1056

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

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J (2004) From the Cover: The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA 101:8066–8071

Nixon JEJ, Wang A, Morrison HG, McArthur AG, Sogin ML, Loftus BJ, Samuelson J (2002) From the Cover: A spliceosomal intron in *Giardia lamblia*. Proc Natl Acad Sci USA 99:3701–3705

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401:877–884

Palmer J, Logsdon JM Jr (1991) The recent origins of introns. Curr Opin Genet Dev 1:470–477

Pawlowski J, Bolivar I, Fahrni JF, de Vargas C, Bowser SS (1999) Molecular evidence that *Reticulomyxa filosa* is a freshwater naked foraminifer. J Eukaryot Microbiol 46:612–617

Pawlowski J, Holzmann M, Berney C, Fahrni JF, Cedhagen T, Bowser SS (2002) Phylogeny of allogromiid Foraminifera inferred from SSU rRNA gene sequences. J Foraminiferal Res 32:334–343

Pawlowski J, Holzmann M, Berney C, Fahrni JF, Gooday AJ, Cedhagen T, Habura A, Bowser SS (2003) The evolution of early Foraminifera. Proc Natl Acad Sci USA 100:11494–11498

Pearson WR (2003) Flexible sequence similarity searching with the FASTA3 program package. Methods Mol Biol 132:185–219

Philippe H, Adoutte A (1998) The molecular phylogeny of eukaryota: solid facts and uncertainties. In: Coombs GH, Vickerman K, Sleigh MA, Warren A (eds) Evolutionary relationships among protozoa. Kluwer Academic, Dordrecht, the Netherlands, pp 25–36

Qiu W-G, Schisler N, Stoltzfus A (2004) The evolutionary gain of spliceosomal introns: sequence and phase preferences. Mol Biol Evol 21:1252–1263

Robertson HM, Warr CG, Carlson JR (2003) Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proc Natl Acad Sci USA 100:14537–14542

Robinson-Rechavi M, Huchon D (2000) RRTree: Relative-rate tests between groups of sequences on a phylogenetic tree. Bioinformatics 16:296–297

Rogozin IB, Wolf YI, Sorokin AV, Mirkin BG, Koonin EV (2003) Remarkable interkingdom conservatiion of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. Curr Biol 13:1512–1517

Ross C, Ross J (1991) Paleozoic Foraminifera. BioSystems 25:39–51

Roy SW, Gilbert W (2005a) Complex early genes. Proc Natl Acad Sci USA 102:1986–1991

Roy SW, Gilbert W (2005b) Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci USA 102:5773–5778

Roy SW, Fedorov A, Gilbert W (2003) Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci USA 100:7158–7162

Sadusky T, Newman AJ, Dibb NJ (2004) Exon junction sequences as cryptic splice sites: implications for intron origin. Curr Biol 14:505–509

Schluter D, Price T, Mooers AO, Ludwig D (1997) Likelihood of ancestor states in adaptive radiation. Evolution 51:1699–1711

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18:502–504

Sheterline P, Clayton J, Sparrow JC (1999) Protein profile. Actin. Oxford University Press, New York

Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116

Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. Proc R Soc Lond B Biol Sci 269:137–142

Sverdlove AV, Rogozin IB, Babenko VN, Koonin EV (2003) Reconstruction of ancestral protosplice sites. Curr Biol 14:1505–1508

Tarrio R, Rodriguez-Trelles F, Ayala FJ (2003) A new *Drosophila* spliceosomal intron position is common in plants. Proc Natl Acad Sci USA 100:6580–6583

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. Syst Biol 51:689–702

Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. Mol Biol Evol 15:1647–1657

Vandekerckhove J, Weber K (1978) At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. J Mol Biol 126:783–802

Wang C-S, Typas MA, Butt TM (2005) Phylogenetic and exon-intron structure analysis of fungal subtilisins: support for a mixed model of intron evolution. J Mol Evol 60:238–246

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699