Widespread Occurrence of Spliceosomal Introns in the rDNA Genes of Ascomycetes

Debashish Bhattacharya,* François Lutzoni,† Valérie Reeb,†‡ Dawn Simon,* John Nason,* and Fernando Fernandez†

*Department of Biological Sciences, University of Iowa; †Department of Botany, Field Museum of Natural History, Chicago, Illinois; and ‡Department of Biological Sciences, University of Illinois at Chicago

Spliceosomal (pre-mRNA) introns have previously been found in eukaryotic protein-coding genes, in the small nuclear RNAs of some fungi, and in the small- and large-subunit ribosomal DNA genes of a limited number of ascomycetes. How the majority of these introns originate remains an open question because few proven cases of recent and pervasive intron origin have been documented. We report here the widespread occurrence of spliceosomal introns (69 introns at 27 different sites) in the small- and large-subunit nuclear-encoded rDNA of lichen-forming and free-living members of the Ascomycota. Our analyses suggest that these spliceosomal introns are of relatively recent origin, i.e., within the Euascomycetes, and have arisen through aberrant reverse-splicing (in *trans*) of free pre-mRNA introns into rRNAs. The spliceosome itself, and not an external agent (e.g., transposable elements, group II introns), may have given rise to these introns. A nonrandom sequence pattern was found at sites flanking the rRNA spliceosomal introns. This pattern (AG-intron-G) closely resembles the proto-splice site (MAG-intron-R) postulated for intron insertions in pre-mRNA genes. The clustered positions of spliceosomal introns on secondary structures suggest that particular rRNA regions are preferred sites for insertion through reverse-splicing.

Introduction

Many eukaryotic genes are interrupted by stretches of noncoding DNA called introns or intervening sequences. Transcription of such "split" genes is followed by a process called RNA splicing, which results in intron removal (Newman 1997). The majority of eukaryotic introns interrupt pre-mRNA in the nucleus and are removed by a ribonucleoprotein complex, termed the spliceosome. Recently, "spliceosomal" introns have also been found in pre-mRNA genes in Chlorella viruses (*pdg* gene; Van Etten and Meints 1999), in genes that encode the small nuclear RNA components of the spliceosome (snRNAs; Tani and Ohshima 1989 [in the ascomycete *Schizosaccharomyces pombe*], 1991; Takahashi et al. 1993 [in the basidiomycetes *Rhodotorula hasegawae* and *Rhodosporidium dacryoidum*]; Biderre, Metenier, and Vivares 1998 [in the microsporidian *Encephalitozoon cuniculi*]), in the small-subunit ribosomal DNA (SSU rDNA) of a limited number of ascomycete fungi (Rogers et al. 1993; Stenroos and DePriest 1998; Myllys, Källersjö, and Tehler 1999; Cubero, Bridge, and Crespo 2000), and in the large-subunit ribosomal DNA (LSU rDNA) of the lichen-forming fungus *Lobaria pulmonaria* (Zoller, Lutzoni, and Scheidegger 1999). The finding of spliceosomal introns in ascomycete rRNA genes is surprising and suggests that splicing factors in these fungi may interact with the nucleolus, the site of ribosome biogenesis (Pederson 1998).

Spliceosomal introns generally contain limited conserved sequences and may reach several thousand nucleotides in length. How these introns spread within and among genes remain central but largely unresolved questions in evolutionary biology (Gilbert 1978; Palmer and Logsdon 1991; Logsdon et al. 1995; Logsdon, Stoltzfus, and Doolittle 1998; Long et al. 1998) because few proven cases of recent and pervasive intron invasion are known (Logsdon 1998). We report here an abundance of spliceosomal introns in both the SSU and the LSU rDNA genes of Euascomycetes fungi. Phylogenetic analyses suggest that these introns are restricted to a monophyletic group within the Euascomycetes and that many have originated relatively recently within this lineage. The Euascomycetes rDNAs provide a concrete example of a recent intron invasion into a family of genes that is otherwise universally free of spliceosomal introns. This study uses the rDNA introns and their flanking exon regions as a model to address fundamental questions about how spliceosomal introns spread and how their sequences evolve.

Materials and Methods
Taxon Sampling

Taxa were selected for this study if their SSU or LSU rDNA sequences were found in preliminary PCR analyses to encode rDNA insertions or if they represented an important lineage of the Ascomycota. Our goal was to present the distribution of the spliceosomal introns within the most complete phylogenetic framework of the Ascomycota we could generate. Taxa were included in this broad phylogenetic analysis only if both targeted portions of the SSU and the LSU rDNA were available. F.L. and V.R. generated 16 of the 40 SSU rDNA sequences and 27 of the 40 LSU rDNA sequences included in the broad phylogenetic analysis of the Ascomycota presented here. The rest of the sequences were retrieved from GenBank (table 1). Our sampling included members of 27 of the 46 orders of Ascomycota as listed in Hawksworth et al. (1995). Six additional LSU rDNA sequences generated by F.F. from closely related

Table 1
Fungal Species Used in this Study, with Voucher Specimen, Culture, or DNA Source Information and GenBank Accession Numbers

Classification	Taxon	SSU Spl. Introns	LSU Spl. Introns	Source for SSU	SSU GenBank No.	Source for LSU	LSU GenBank No.
Ascomycota (AS)							
Euascomycetes (EU)							
Pyrenomycetes (P)							
Halosphaeriales	Aniptodera chesapeakensis			SMH 3327 (culture)[a]	AF279373	SMH 3327 (culture)	AF279374
Sordariales	Chaetosphaeria sp. 17		678	—	—	SMH 1725 (culture)	AF279416
	Chaetosphaeria sp. 17		678	—	—	SMH 3854 (culture)	AF279417
	Chaetosphaeria sp. 7		678	—	—	SMH 1267 (culture)	AF279418
	Chaetosphaeria sp. 8		678	—	—	SMH 2893 (culture)	AF279419
	Melanochaeta sp. 2		678	—	—	SMH 1971 (culture)	AF279420
	Melanochaeta sp. 8		678, 784	—	—	SMH 3802 (culture)	AF279421
Diaporthales	Diaporthe phaseolorum			—	L36985	—	U47830
Microascales	Ceratocystis fimbriata			—	U43777	—	U17401
Phyllachorales	Glomerella cingulata			—	U48427	—	U17403
Hypocreales	Hypocrea schweinitzii			—	L36986	ATCC 90178[b]	AF279395
Melanommatales?	Mycopepon smithii			SMH 1609 (culture)	AF279399	SMH 1609 (culture)	AF279400
Xylariales	Seynesia erumpens			SMH 1291 (culture)	AF279409	SMH 1291 (culture)	AF279410
Loculoascomycetes							
Dothideales	Curvularia brachyspora			—	L36995	ATCC 58872	AF279380
Arthoniales (A)	Dendrographa minor			R. Ornduff 10,070 (DUKE)[c]	AF279381	R. Ornduff 10,070 (DUKE)	AF279382
	Arthonia lapidicola	1510		Grube, Gargas, and DePriest (1996)	NA		—
	Roccella spp.	300, 330, 393		Myllys, Källersjö, and Tehler (1999)	AF110340 AF110343 AF110346		—
	Schismatomma pericleum				U23540	Tehler 7701 (S)	AF279408
Inoperculate discomycetes (I)							
Helotiales	Cudonia circinans			—	AF107343	JP 232 (DNA)[d]	AF279379
	Spathularia velutipes			—	AF281072	JP 173 (DNA)	AF279411
"Lecanoromycetes" (L)							
Patellariales	Arthrorhaphis citrinella	265		Lutzoni 97.06.29 ST2. 1 (DUKE)	AF279375	Lutzoni 97.06.29 ST2.1 (DUKE)	AF279376
Gyalectales	Dimerella lutea	882, 939	1091	Ryan 31430 (ASU)	AF279386	Ryan 31430 (ASU)	AF279387
Ostropales	Diploschistes scruposus	883		Reeb 12-X-97/10 st 4.1 (DUKE)	AF279388	Reeb 12-X-97/10 st 4.1 (DUKE)	AF279389
	Graphis scripta	1416		Winka, Ahlberg, and Eriksson (1998)	AF038878	—	
Gyalectales	Gyalecta jenensis	265, 393	678, 711, 978	Lutzoni 98.08.17-6 (DUKE)	AF279390	Lutzoni 98.08.17-6 (DUKE)	AF279391
Lecanorales	Hymenelia epulotica			Lutzoni 96.08.29-2b st.4.2 (DUKE)	AF279393	Lutzoni 96.08.29-2b st.4.2 (DUKE)	AF279394
	Trapeliopsis granulosa		858	Lumbsch and Feige 10.7.1994 (DUKE)	AF279414	Lumbsch and Feige 30.6.1992 (DUKE)	AF279415
	Dibaeis baeomyces		775, 787		AF113712	Lutzoni, 93.08.20-1 1/1 (DUKE)	AF279385
Pertusariales	Ochrolechia frigida	330		Lutzoni 93.07.22-2 (DUKE)	AF279402	Lutzoni 93.07.22-2 (DUKE)	AF279403
	Pertusaria trachythallina	330		Stenroos and DePriest (1998)	U64685		—
Lecanorales	Stereocaulon paschale	330		Reeb STEPAS (DUKE)	AF279412	Reeb STEPAS (DUKE)	AF279413
	Sterocaulon spp.	330		Stenroos and DePriest (1998)	U64687-8		

Table 1
Continued

Classification	Taxon	SSU Spl. Introns	LSU Spl. Introns	Source for SSU	SSU GenBank No.	Source for LSU	LSU GenBank No.
	Cladonia spp.	330		Stenroos and DePriest (1998)	U64682–3	—	—
	Cladina portentosa	330		Stenroos and DePriest (1998)	U64684	—	—
	Physcia aipolia	330		Stenroos and DePriest (1998)	U64686	—	—
	Physconia spp.	296, 297, 331, 513, 673, 943, 1129		Cubero, Bridge, and Crespo (2000)[e]	NA	—	—
Trichotheliales	*Porina guentheri*		**678**	**Lutzoni 97.10.09-10 2/2 st 1 (DUKE)**	**AF279404**	**Lutzoni 97.10.09-10 2/2 st 1 (DUKE)**	**AF279405**
Peltigerales	*Lobaria pulmonaria*	**393**	678		AF18935		AF183934
	Lobaria quercizans		678	**17-III-1992 (DUKE)**	AF279396	**17-III-1992 (DUKE)**	AF279397
Chaetothyriales	*Capronia pilosella*		776		U42473	**SMH 2565 (culture)**	AF279378
	Phialophora americana	1510		Rogers et al. (1993)	X65199		—
Verrucariales	*Dermatocarpon americanum*	**390**		**Merrill 0014313**	**AF279383**	**Merrill 0014313 (DUKE)**	**AF279384**
Pyrenulales	*Pyrenula cruenta*	**332, 393**		**Lutzoni 98.06.17-4 (DUKE)**	**AF279406**	**Lutzoni 98.06.17-4 (DUKE)**	**AF279407**
Eurotiales (E)	*Chromocleista malachitea*				D88323		AB000621
	Hamigera avellanea				D14406		AB000620
Operculate discomycetes (O)							
Pezizales	*Morchella esculenta*				U42642	**ATCC 10968**	**AF279398**
	Otidea leporina				U53381	—	Z49795 and Z49791
	Urnula hiemalis				Z49754	—	Z48319 and Z48320
Hemiascomycetes (H)							
Saccharomycetales	*Candida albicans*				X53497	—	X70659
	Hormoascus platypodis				L36984	**ATCC 24747**	**AF279392**
	Saccharomyces cerevisiae				J01353	—	J01355
Archiascomycetes							
Neolectales	*Neolecta irregularis*				Z47721	**Jamie Platt JP176 (DUKE)**	**AF279401**
Schizosaccharomycetales	*Schizosaccharomyces pombe*				X54866	—	Z19136
Taphrinales	*Taphrina pruni*				AB000956	—	Z49792 and Z49793
Basidiomycota (outgroup)	*Athelia bombacina*				M55638	**ATCC 20629**	**AF279377**
	Coprinus cinereus				M92991		AF041494

NOTE.—Insertion sites of new spliceosomal (spl.) introns determined in this study are shown in boldface. New sequences (with sources of DNA) generated by F.L. and V.R. or by F.F. that were deposited in GenBank are shown in boldface. "NA" indicates that sequences are not available in GenBank.

[a] SMH corresponds to fungal cultures provided by Sabine Huhndorf (Department of Botany, Field Museum of Natural History, Chicago).

[b] ATCC corresponds to fungal cultures obtained from the American type Culture Collection.

[c] When the name of the collector, the collection number, and the herbarium acronym (in parentheses) are provided, this means that DNA was from herbarium specimens or fresh material deposited at various herbaria.

[d] DNA corresponds to DNA aliquots provided by Jamie Platt (Department of Plant and Microbial Biology, University of California, Berkeley).

[e] The 24 introns in the SSU rRNA gene of *Physconia* spp. are not individually listed here and can be found in table 1 in Cubero, Bridge, and Crespo (2000).

members of the genera Chaetosphaeria and Melano-chaeta were included in a separate phylogenetic analysis of the 678 spliceosomal intron. The 49 new sequences presented here were deposited in GenBank (table 1).

DNA Isolation and Sequencing

Genomic DNA was obtained from fresh samples, herbarium specimens, fungal cultures, or DNA aliquots sent to us from other labs (table 1). Except for the latter source, DNA was isolated using the Puregene Kit (GEN-TRA Systems) following the manufacturer's protocol for filamentous fungi. Genomic DNA was examined for quality and quantity on an ethidium-bromide–stained TBE 1% agarose gel. Symmetric polymerase chain re-action (PCR) was performed on three different concen-trations of DNA to amplify targeted 1.0- and 1.4-kb fragments at the 5′ end of the SSU and LSU rDNA genes, respectively. Details of the different reaction con-ditions used in the PCR amplifications are available on request from F.L. The PCRs were done with the follow-ing primer pairs: (1) SSU rDNA—nSSU131-NS22, nSSU97a-NS22, or nSSU97b-NS22 (Gargas and Taylor 1992; unpublished data); (2) LSU rDNA—LROR-LR7, LIC15R-LR7, LIC24R-LR7, LROR-LIC2044, LIC15R-LIC2044, LIC24R-LIC2044, LROR-LIC2028, LIC15R-LIC2028, or LIC24R-LIC2028 (Vilgalys and Hester 1990; Rehner and Samuels 1994; Miadlikowska and Lutzoni 2000; unpublished data). The PCR products were purified using GELase Agarose Gel-Digesting Preparation (Epicentre Technologies) following the manufacturer's instructions. Both strands of the purified PCR products were sequenced using the following prim-ers: (1) SSU rDNA—nSSU97a, nSSU97b, nSSU131, SR11R, SR7, SR7R, nSSU634, nSSU897R, nSSU1088, nSSU1088R, and NS22 (Gargas and Taylor 1992; Spa-tafora, Mitchell, and Vilgalys 1995; unpublished data; http://www.botany.duke.edu/fungi/mycolab); (2) LSU rDNA—LROR, LIC15R, LIC24R, LIC52R, LR3, LR3R, LR5, LR5R, LR6, LR6R, LIC2028, LIC2044, and LR7 (Vilgalys and Hester 1990; Rehner and Sa-muels 1994; Miadlikowska and Lutzoni 2000; http://www.botany.duke.edu/ fungi/mycolab). The sequencing reaction was performed in a 10-μl final volume using dRhodamine Terminator (ABI PRISM, Perkin-Elmer Biosystems), Thermo Sequenase dye terminator (Amer-sham), or BigDye Terminator (ABI PRISM, Perkin-El-mer Biosystems) following the manufacturer's instruc-tions. Sequenced products were precipitated with 10 μl of deionized sterile water, 2 μl of 3 M NaOAC, and 50 μl of 95% EtOH. Polyacrylamide gel electrophoresis was conducted using Long Ranger Singel packs (FMC BioProducts) and an ABI 377A automated DNA se-quencer (Perkin-Elmer, Applied Biosystems). Each se-quence fragment was subjected to a blast search to ver-ify its identity. Sequence fragments were assembled us-ing Sequencher 3.0 (Gene Codes).

RT-PCR Experiment

RT-PCR analyses were done (as in Bhattacharya, Stickel, and Sogin 1993) with Stereocaulon paschale and *Lobaria quercizans* to test for intron presence/ab-sence in mature rRNAs. For Stereocaulon, cDNA syn-thesis was initiated with a primer complementary to the 3′ terminus of the SSU rRNA (Medlin et al. 1988), and PCR was done with two primers that recognized sites flanking the intron at position 330 (numbering based on the *Escherichia coli* gene) in the SSU rRNA. The 5′ primer was 160 nt upstream of the intron (5′-GGTGATTCATAATAACTCAACG-3′), whereas the 3′ primer was 654 nt downstream (5′-ACACCGTCCGA-TCCCCAGTCGG-3′). For Lobaria, cDNA synthesis was initiated with a primer complementary to a region near the 3′ terminus of the LSU rRNA (5′-TTCATTCGGCCGGTGAGTTG-3′), and PCR was done with primers that flanked the spliceosomal intron at position 678. The 5′ PCR primer was 48 nt upstream of the 678 intron (5′-GCACCATCGACCGATCCTGA-3′), whereas the 3′ primer was 210 nt downstream (5′-TAGGTTAAGGCTGTTTCAGC-3′) of this intron.

Alignment and Phylogenetic Analyses

All alignments were generated using the program Sequencher, version 3.0 (Gene Codes), and manually op-timized. With the exception of *Arthrorhaphis citrinella,* if phylogenetic analyses revealed an unexpected result, a second sequence was obtained from a distant population of the same species or from a closely related species. If these two sequences were not found to be sister to each other in the phylogenetic tree, additional sequences were obtained until two of them would form a monophyletic entity. The remaining orphan sequences were discarded.

The combined alignment included 40 species and 6,494 characters. A total of 5,692 sites were excluded from the phylogenetic analysis. These sites included constant characters, all introns, and regions that were ambiguously aligned due to the presence of gaps. Of the remaining 802 characters, 516 were parsimony-infor-mative. The unambiguously aligned portions of the SSU and LSU rDNA genes were each subjected to a specific symmetric step-matrix, taking into account the estimated empirical frequency of all changes (i.e., all six substi-tution types and four single-position indel types) for these regions as described in Lutzoni (1997). The rare, unambiguously aligned gaps were treated as a fifth char-acter state. Ambiguously aligned portions of the align-ment that included no more than 15 different sequences were coded for a maximum of 15 character states per character using the program INAASE, version 0.2c1 (Lutzoni, et al. 2000). A total of 16 ambiguous regions fit that criterion, forming 16 additional characters, for a total of 6,510 characters, 532 of which were parsimony-informative. Each of these 16 coded characters was sub-jected to a specific symmetric step-matrix accounting for the optimal number of changes between all possible combinations of two sequences found within that am-biguous region. The combined phylogenetic analysis of these 40 species was performed using maximum parsi-mony as the optimization criterion with 1,000 random-addition sequences, TBR swapping, and MULTREES selected. Bootstrap support (Felsenstein 1985) was es-

timated by 1,000 replicates, implementing full heuristic searches with five random-addition sequences per replicate. All phylogenetic analyses were performed with PAUP* (Swofford 1999).

Tests of the Proto-Splice Site Hypothesis

We used two statistical tests to determine whether sequences flanking Euascomycetes spliceosomal introns showed a nonrandom base composition. In the first method, we used a likelihood ratio test for multinomial data (Sokal and Rohlf 1995) to assess the goodness of fit of the MAG—R motif hypothesis (hereinafter, the intron position is indicated with ''—'') to the flanking sequence data observed for 27 events of spliceosomal intron insertion into Euascomycetes rDNA. The null hypothesis specified that nucleotide usage at the four proto-splice site nucleotides was random and depended on the nucleotide composition of Euascomycetes SSU and LSU rDNA sequences in general. These expected nucleotide probabilities were estimated from the observed nucleotide frequencies over all sites for 80 rDNA sequences (A = 26%, C = 22%, G = 27%, T = 25%). The alternative hypothesis specified that the null hypothesis was not true and allowed the nucleotide probabilities to be free of the constraint that they sum to 1. These probabilities were determined from the observed frequencies of nucleotides at each of the four flanking sites for 27 intron insertion events. The likelihood ratio test statistic ($-2 \log \Lambda$) was determined for each flanking site (one degree of freedom each) and then summed over sites to test the MAG—R hypothesis. The large sample distribution of the cumulative test was a chi-square distribution with four degrees of freedom. Because the expected counts for the four possible nucleotides at each of the four different positions were relatively small (i.e., less than 7), we also simulated nucleotide frequencies in the motif (1,000 trials) under the null hypothesis of random base usage. The cumulative likelihood ratio test statistic was determined for each simulated data set in order to construct its probability distribution under the null model.

In the second examination of flanking-sequence composition, we calculated the Pearson's chi-square statistic for each of the four proto-splice site nucleotides to determine whether individual nucleotide positions at these sites showed a significant departure from the null expectation of random base usage. For data obtained from 27 spliceosomal intron insertion events, one cell in the chi-square test was for the count of nucleotides predicted by the MAG—R hypothesis (e.g., G) and the other was for all other nucleotides (e.g., A/C/T). The significance of each of these tests was determined using one degree of freedom.

Results and Discussion
The Short rDNA Insertions Are Putatively Spliceosomal Introns

Comparison of the SSU rDNA sequences resulted in an alignment of total length 3,325 nt, due primarily to the presence of eight spliceosomal and five group I introns at different genic sites. The alignment for the sequenced region of the LSU rDNA gene was of length 3,169 nt. The length discrepancy between this sequenced region and the resulting alignment was also due primarily to the presence of eight spliceosomal and five group I introns at different genic sites. Of the 27 spliceosomal intron insertion sites that were available at the time of this study, 14 novel positions come from our sequence analyses. These are the 265, 332, 390, 882, 883, and 939 introns in the SSU rRNA gene and the 711, 775, 776, 784, 787, 858, 978, and 1091 introns in the LSU rRNA (see table 1). Previous analyses have shown insertions at the 300, 330, 393, 1510 (Myllys, Källersjö, and Tehler 1999), 1416 (Winka, Ahlberg, and Eriksson 1998), 296, 297, 331, 513, 673, 943, and 1129 SSU rRNA sites (Cubero, Bridge, and Crespo 2000) and at the 678 LSU rRNA site (Zoller, Lutzoni, and Scheidegger 1999) of different fungi to be spliceosomal introns. These sites are numbered according to their homologous positions in the *E. coli* rDNA genes. The 330 intron in *Physconia* spp. (Cubero, Bridge, and Crespo 2000) was repositioned in our analyses at position 331, making this a novel site in the SSU rRNA gene. We analyzed a total of 69 introns at the 27 different insertion sites; 29 introns came from our study, 14 came from Myllys, Källersjö, and Tehler (1999), the *Graphis scripta* 1416 SSU rRNA intron came from Winka, Ahlberg, and Eriksson (1998), 24 introns in Physconia species came from Cubero, Bridge, and Crespo (2000), and the 678 LSU rRNA intron in *L. pulmonaria* came from Zoller, Lutzoni, and Scheidegger (1999). The 393 intron in *Umbilicaria umbilicarioides* reported in Myllys, Källersjö, and Tehler (1999) was not available from GenBank. The taxa containing the 29 introns that we found are identified in table 1 and figure 1. Chaetosphaeria and Melanochaeta strains/species contained seven introns at the 678 and 784 LSU rRNA sites. These taxa were not included in the phylogenetic analyses because only the LSU rRNA region was determined.

All of these rRNA insertions were of short length (49–199 bp) and contained the conserved spliceosomal intron donor (5′-GUAAGU-3′) and acceptor (5′-YAG-3′) sites (fig. 2A). The intron donor and acceptor sites were 5′-GUAUGU-3′ and 5′-YAG-3′, respectively, for yeast mRNA introns (Rymond et al. 1990; Lopez and Séraphin 1999). In addition, the highly conserved branch site (5′-UACUAAC-3′ Lopez and Séraphin 1999) that interacts with the U2 snRNA was nearly perfectly conserved (5′-URCUAAC-3′) in the rDNA introns. A lack of available Euascomycetes pre-mRNA gene sequences prevented, however, a direct comparison of donor, acceptor, and branch site regions in rRNA and mRNA introns in these taxa. Apart from these regions, all other rRNA intron sequences showed no apparent pattern of sequence conservation (see fig. 2B). This was highlighted by the putatively vertically transmitted intron found at position 678 in the Pyrenomycetes. High divergence within the Pyrenomycetes 678 introns allowed only the reliable alignment (fig. 2B) of sequences from members of the same species (e.g., *Chaetosphaeria* sp. 17; 85% identity), whereas the *Chaetosphaeria* spp.
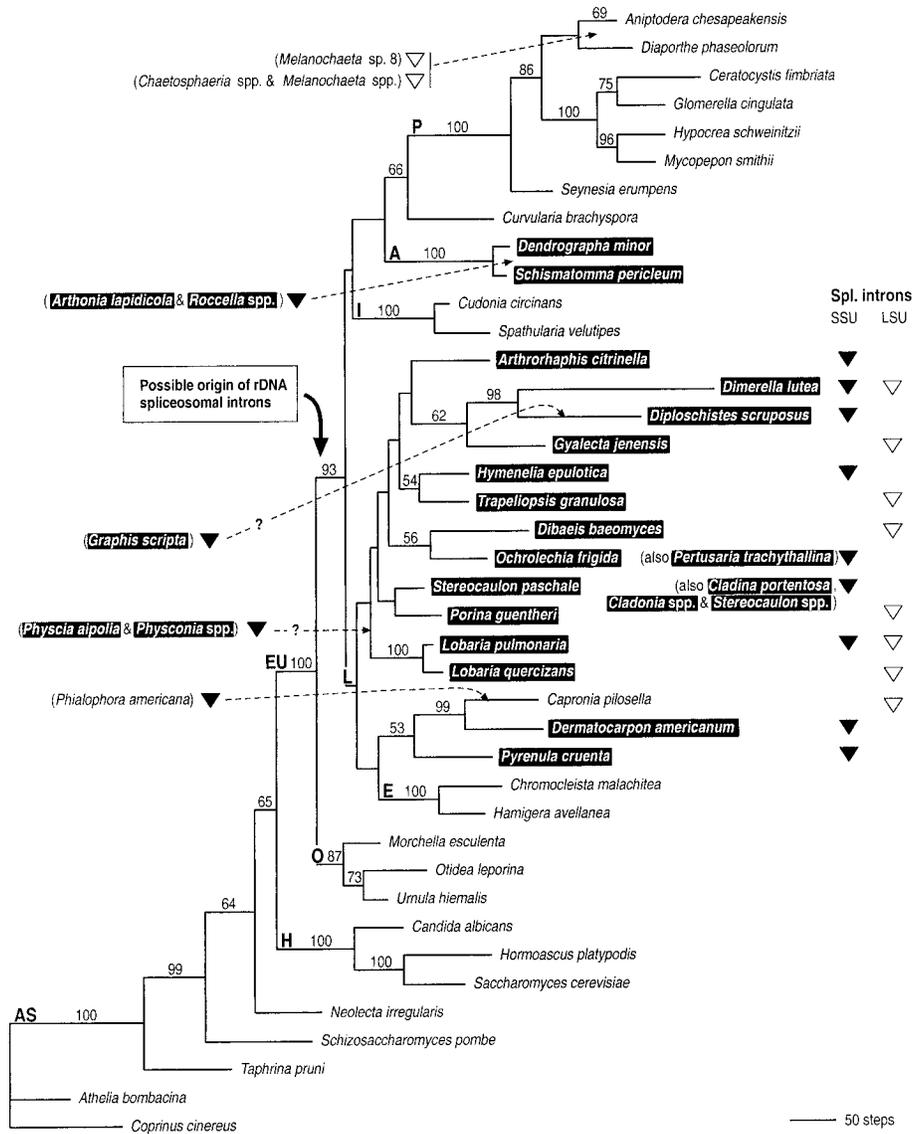
FIG. 1.—Phylogenetic relationships among major groups of ascomycetes based on an analysis of SSU and LSU rDNA sequences from 40 species representing 26 of the 46 orders of Ascomycota as listed in Hawksworth et al. (1995). This is the single most-parsimonious tree (4,336.60 steps) revealed by weighted maximum parsimony (PAUP*; Swofford 1999). This topology was recovered 399 times out of the 1,000 random addition sequences using the tree bisection-reconnection (TBR) branch-swapping algorithm (consistency index = 0.51; consistency index excluding uninformative characters = 0.42; retention index = 0.41). Internodes of maximum length 0 were collapsed to form polytomies. The ascomycetes were outgroup-rooted with two basidiomycete species. Bootstrap values >50% are shown for each internode. Fungal species shown in white on a black background are obligate lichen symbioses. The monophyletic origins of the Ascomycota (AS), Hemiascomycetes (H), Euascomycetes (EU), operculate discomycetes (O), inoperculate discomycetes (I), Eurotiales (E), Pyrenomycetes (P), and Arthoniales (A) are highly supported, with bootstrap values of >85% (see also table 1). The "Lecanoromycetes" (L) form a monophyletic but poorly supported group. The distribution of spliceosomal introns in the SSU and LSU rDNA genes are shown with inverted full and empty triangles, respectively. Names of taxa containing spliceosomal introns which could not be included in the phylogenetic analysis due to a lack of either SSU or LSU rDNA sequences have simply been included (in parentheses) in the figure on either side of the tree. The approximate positions of these taxa were marked with arrows if they did not share an intron site with a closely related taxon included in the phylogenetic analysis. The phylogenetic positions of Graphis, Physcia, and Physconia are not clear. The possible origin of rDNA spliceosomal introns in the Euascomycetes is shown on the left side of the tree.

and *Melanochaeta* spp. introns did not share significant sequence identity and could not be aligned outside of the donor, acceptor, and branch sites (38%–57% overall identity). Comparison of partial (987 nt) coding sequences of the LSU rDNA from these taxa showed 99% identity between these two members of *Chaetosphaeria* sp. 17 and 94%–96% identity between *Chaetosphaeria* spp. and *Melanochaeta* spp. (results not shown). This evidence for a pattern of sequence conservation typical of spliceosomal introns is important because an alternative explanation for the origin of these Euascomycetes rRNA spliceosomal introns is that they are highly reduced or "degenerate" group I introns (Stenroos and DePriest 1998).

A

	+1	+2	+3	+4	+5	+6									+3	+2	+1
A	0	0	50	46	2	3	10	24	1	0	68	69	1	1	69	0
C	0	3	0	12	0	19	9	0	67	0	0	0	63	20	0	0
G	69	0	9	0	67	0	4	43	0	0	1	0	1	1	0	69
U	0	66	10	11	0	47	46	2	1	69	0	0	4	47	0	0
	G	**U**	**A**	**A**	**G**	**U**	**U**	**R**	**C**	**U**	**A**	**A**	**C**	**Y**	**A**	**G**
	100%	96%	72%	67%	97%	68%			97%	100%	99%	100%	91%	97%	100%	100%

B

```
                              donor site                                                              branch site      acceptor site
C.sp.17  (1725)-678  GAAAGgugagcacuaaaaucuucauuauagaaaggcagua------------------------------------------uacuaac---gauuaucc-cagAUGGU
C.sp.17  (3854)-678  GAAAGgugagaaucaauucuucauuauaguaaggcagua------------------------------------------ugcuaac---aauugucccuagAUGGU
C.sp.7   (1267)-678  GAAAGguaagcaccaacccccgauccugcucc-gc-gggaggcccg-----------------------------------uacuaac----agugu-ggcagAUGGU
C.sp.8   (2893)-678  GAAAGguaagcaccuaucucuc---cccgcucccgccggggccuaug----------------------------------uacuaac----ugugu-gacagAUGGU
Mel.sp.2 (1971)-678  GAAAGgugagcaagaucuauuuucuuuuuuagccuaugccagggcgcuuaugguugugaaauuuauaucuugaagcuauugcuaacagaaucucu-uguagAUGGU
Mel.sp.8 (3802)-678  GAAAGgugagaaacauuuuuucccacuuugcaucuguuuuuuaaaguuugac-----------------------------ugcuaac----uuucu-gaaagAUGGU
```

Fig. 2.—Analysis of rDNA spliceosomal introns in the Euascomycetes. *A,* Sequence conservation of the intron donor, acceptor, and branch sites in 69 different introns at 27 rDNA sites in the Ascomycetes. Twenty-nine introns were found in this study; those remaining came from Rogers et al. (1993), Stenroos and DePriest (1998), Winka, Ahlberg, and Eriksson (1998), Myllys, Källersjö, and Tehler (1999), Zoller, Lutzoni, and Scheidegger (1999), and Cubero, Bridge, and Crespo (2000). *B,* Alignment of the homologous 678 spliceosomal intron in the Pyrenomycetes. Note the high intron sequence identity only between members of the same species (i.e., *Chaetosphaeria* sp. 17). The intron donor, acceptor, and branch sites have been marked.

Fungal rDNA genes are rich in group I introns, some of which have been identified as putative degenerate forms (Gargas, DePriest, and Taylor 1995; Grube, Gargas, and DePriest 1996; Myllys, Källersjö, and Tehler 1999). We do not believe that the spliceosomal introns are degenerate group I introns for four major reasons (see also Cubero, Bridge, and Crespo 2000). First, they share no primary sequence or secondary structure similarity with any known group I introns. Second, they contain the clear sequence landmarks of spliceosomal introns, such as the GU—AG consensus sequence at the intron termini and the branch site sequence (see above). Third, a degenerate intron is primarily identified as such (in the absence of significant sequence identity with group I introns) because it is at an rDNA position that harbors "normal" group I introns in other fungi and in other eukaryotes (Gargas, DePriest, and Taylor 1995; Grube, Gargas, and DePriest 1996). The 69 rDNA spliceosomal introns that have been summarized in figure 2A are all found at sites (except for the 943 insertions) that do not contain group I introns in eukaryotes. The 943 spliceosomal introns are typical pre-mRNA type insertions which do not appear to trace their ancestry to group I introns at this site (Cubero, Bridge, and Crespo 2000). Fourth, and most importantly, only 2 of 27 of the spliceosomal intron sites (the 882 intron in the SSU rDNA gene of *Dimerella lutea* and the 296 intron in *Physconia enteroxantha*) are preceded by a U at the 5′ intron-exon junction. The U forms a nearly perfectly conserved U·g base pair with the intron-encoded g in folding segment P1 that marks the 5′ splice site of group I introns (Cech 1988).

To test whether the spliceosomal introns are present in the mature transcript, PCR was done with cDNA copies of rRNA from the lichen-forming fungi *S. paschale* and *L. quercizans.* Comparison of the fragments resulting from this experiment showed that the introns were not present in the mature rRNA sequences (fig. 3). Cubero, Bridge, and Crespo (2000) also used the RT-PCR method to show that multiple spliceosomal introns were

Stereocaulon paschale   Lobaria quercizans

Fig. 3.—Photograph of agarose gel showing the results of RT-PCR analyses using total RNA from *Stereocaulon paschale* and *Lobaria quercizans.* The size shift of the bands from the genomic DNA (gDNA) to the complementary (cDNA) indicates that the 330 spliceosomal intron has been spliced out of the mature SSU rRNA of Stereocaulon and the 678 intron has been excised from the Lobaria mature LSU rRNA.

not present in the mature SSU rRNA of Physconia species. The cDNA fragments from Stereocaulon and Lobaria were sequenced over the intron region and compared with the homologous genomic region. This analysis confirmed the 330 intron insertion site in Stereocaulon and the 678 intron site in Lobaria. In addition, we did Southern blot analysis of genomic DNA from Stereocaulon using a probe fragment of size 930 bp that encoded the 330 intron (70 bp) and flanking SSU rDNA sequence. The results of this analysis showed single fragments in *Eco*RI- and *Bam*HI-digested DNAs (results not shown), consistent with the idea that the 330 intron is restricted to the SSU rDNA gene and is not a member of a mobile family of transposable elements. We recognize, however, that sites of intron insertion, evidence of intron excision from mature rRNA, and intron distribution within the nuclear genome need to be verified for each intron in different species. Future studies in our labs will address these issues and test the alternative hypotheses that some insertions remain in the mature rRNA and that some rRNA introns may not be limited to these loci in Euascomycetes. The finding of nonspliced spliceosomal intron-like insertions would be of great interest. Of particular relevance is the case in which phylogenetic analyses show that an intron which is excised in all ancestors is no longer recognized by the spliceosome in a descendant and remains in the mature rRNA. Such inactivated intron sequences will provide important insights into the evolution of splicing signals. We predict, however, that virtually all spliceosomal introns will be excised from the pre-rRNA, because these introns generally interrupt highly conserved regions of the coding regions (e.g., the sites of mRNA and tRNA interaction; Green and Noller 1997).

On the basis of this evidence, we hypothesize that the multiple short insertions that were identified in the rDNA genes of the Euascomycetes are spliceosomal introns. These sequences are of recent origin (i.e., they are restricted to a subgroup of the Ascomycota [see fig. 1]) and are therefore well suited for studying the origin and evolution of spliceosomal introns. Unequivocal proof that these insertions are spliceosomal introns, however, requires in vivo evidence of splicing mediated by the spliceosome.

Spliceosomal Intron Origin

Spliceosomal introns are believed to originate primarily through two processes: transposable element/group II intron insertion or duplication of a preexisting intron (Palmer and Logsdon 1991; Purugganan and Wessler 1992; Logsdon, Stoltzfus, and Doolittle 1998; Nouaud et al. 1999). Although transposable elements can sometimes insert so that they behave like spliceosomal introns (e.g., Giroux et al. 1994), they generally result in the introduction of insertions or deletions (indels) of sequences at the insertion site (Patthy 1996). Analyses of intron-flanking sequences do not support the idea that indels are normally associated with intron insertion into protein-coding regions (e.g., Dibb and Newman 1989; Weber and Kabsch 1994; Bhattacharya

and Weber 1997). In addition, recent analyses underline the extensive genetic changes (e.g., origin of a new promoter, incorporation of genomic sequences) that are required for the origin of an intron within a "domesticated" P element insertion in Drosophila species (Nouaud et al. 1999). This type of data argues against a general transposon model of intron spread. In the group II intron scenario, spliceosomal introns originate from autocatalytic group II introns located in organellar genomes that have been transferred to the nucleus and thereafter inserted themselves into nuclear genes through endonuclease-dependent reverse-splicing (retrohoming; Saldanha et al. 1993; Belfort and Perlman 1995; Zimmerly et al. 1995; Eskes et al. 1997; Cousineau et al. 1998). Recently, a group II intron has been found that can reverse-splice into RNA in an endonuclease-independent manner (Cousineau et al. 2000). This latter mechanism is the same as that first proposed for group I intron spread (Cech 1985; Sharp 1985; see below) and can lead to intron insertion into heterologous RNA sites in the sense strand. In both cases of group II intron insertion, the autocatalytic capacity of the intron allows its removal (albeit inefficiently; Cousineau et al. 2000) from precursor RNA, and over time, the insertions are presumptively reduced to spliceosomal introns (Cavalier-Smith 1991; Perlman and Podar 1996). The demonstration of endonuclease-independent reverse-splicing offers the first direct link between group II and spliceosomal intron origin. This finding also suggests that reverse-splicing may be an important mechanism of intron spread (Eickbush 2000).

The alternative theory for the "recent" (i.e., not implicated in the assembly of ancient genes; see de Souza et al. [1998] and Roy et al. [1999] for details) origin of spliceosomal introns postulates that introns are spread via duplication of existing sequences. Reverse trans-splicing is thought to facilitate insertion of existing introns into RNAs with the splicing machinery itself responsible for the spread of these sequences into new sites both within and between transcripts (Sharp 1985; Fink 1987; Martinez, Martin, and Cerff 1989). After reverse-splicing, reverse transcription and homologous recombination of the intron-containing cDNA with the genomic copy would result in intron lateral transfer. An appealing aspect of the reverse-splicing model is that it facilitates the spread of intact introns containing the conserved donor, acceptor, and branch sites into novel RNA positions without the need for these regions to evolve into the sequences required for excision. In addition, if introns are reverse-spliced into exon sequences which have a high affinity for splicing factors ("proto-splice" sites; Dibb and Newman 1989), then they would be efficiently recognized in the transcript that contains the laterally transferred intron (Stephens and Schneider 1992). Providing a mechanism for "clean" intron insertion, as would result from reverse-splicing of existing spliceosomal introns, is important because many introns interrupt highly conserved genes which may exist in single copies (Dibb and Newman 1989; Bhattacharya and Weber 1997). Imperfect insertion events could wreak havoc in single-copy genes because the introns would

A

Panel A: 5' Exon — intron insertion — 3' Exon

	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6	+7	+8	+9
A	3	5	4	6	11	10	9	12	2	5	3	4	7	10	10	16	8	10
C	6	3	7	7	3	3	8	5	3	2	6	3	3	7	6	8	6	6
G	9	15	14	12	9	5	4	3	20	18	12	12	9	6	7	2	6	4
U	9	4	2	2	4	9	6	7	2	2	6	8	8	4	4	1	7	7
	N	G	G	N	N	N	N	A	G	G	N	N	N	N	N	N	A	N

(boxed: A 44% G 74% G 67%)

B

Distribution of likelihood ratio test statistics determined by simulation (n=1000) under the null hypothesis.

Actual test value (−2 log Λ) = 22.15 for MAG-R motif p<0.005 (df=4)

(y-axis: Probability, 0.00–0.25; x-axis: Likelihood ratio test statistic, 0–22)

C

Motif:	A/C (vs G/U)	A (vs C/G/U)	G (vs A/C/U)	A/G (vs C/U)
Position	-3	-2	-1	+1
A	9	12	2	5
C	8	5	3	2
G	4	3	20	18
U	6	7	2	2
chi-square	2.42	4.77	28.05	11.23
df=1	n.s.	P<0.05	P<0.001	P<0.001

FIG. 4.—Sites of rRNA intron insertion. *A,* Conservation of flanking sequences surrounding 27 different spliceosomal intron insertion sites in the rDNA genes of Euascomycetes. The putative proto-splice site region is boxed. Nucleotides outside of the proto-splice site are identified when they occur in ≥14 of 27 of the sequences. *B,* Likelihood ratio test of the hypothesis that nucleotide composition in four sites surrounding rRNA intron insertions is consistent with random usage given Euascomycetes rDNA nucleotide frequencies. The alternative hypothesis is the motif MAG—R. *C,* Chi-square test to determine whether the individual nucleotides in the MAG—R proto-splice site differ from the null expectation of random base usage in Euascomycetes rRNAs.

most certainly result in loss-of-function mutations, possibly leading to grave consequences for the organism (most obviously in haploids). The reverse-splicing (i.e., mediated by the spliceosome) model therefore provides an explanation for intron spread that does not depend on external agents such as transposable elements/group II introns or the requirement for insertions to evolve into intron-like sequences to drive frequent and recent intron lateral transfers. Given that reverse-splicing may be an important mechanism for rRNA spliceosomal intron origin, are there preferred sites for intron insertion?

Testing the Proto-Splice Site Hypothesis

The proto-splice site hypothesis posits that there are target sequences (e.g., MAG—R; Dibb and Newman 1989) that are preferred sites for intron insertion. The proto-splice site (at least a ''general'' proto-splice site) hypothesis, however, was not unequivocally supported by a recent analysis of intron-flanking sequences in the pre-mRNA genes of six model eukaryotes (Long et al. 1998; but see Logsdon 1998). These analyses show that introns are not evenly distributed over codon positions (phase 0 introns are favored; Tomita, Shimuzu, and Brutlag 1996) and fail to show the existence of a symmetric intron distribution with respect to codon structure, a prediction of the proto-splice site hypothesis (Long et al. 1998). Given these conflicting ideas, we studied the sequences flanking the rDNA spliceosomal introns to determine if they encoded a conserved target sequence for intron insertion (fig. 4*A*). It is important to note that the proto-splice site need not be perfectly conserved in dif-

ferent organisms but rather is a set of nucleotides that, with some statistical uncertainty, shows a nonrandom sequence pattern at sites flanking introns. It is conceivable that proto-splice sites may differ between lineages, reflecting, for example, differences in how spliceosomes recognize introns (e.g., exon definition hypothesis; Berget 1995; McCullough and Berget 1997). If this is true, Euascomycetes may have a pattern different from that found in animals or other fungi.

Keeping these ideas in mind, we used two different statistical methods to test for the presence of a proto-splice site at sites of spliceosomal intron insertion in Euascomycetes rRNA genes. In the first method, we used the likelihood ratio test for goodness of fit to determine whether the MAG—R motif (Dibb and Newman 1989) in pre-mRNA genes is conserved in rRNAs (fig. 4*B*), and second, we used a chi-square test to determine whether individual nucleotides in the rRNA proto-splice site showed a significant departure from the null expectation of random base usage (fig. 4*C*). The likelihood ratio test showed significant support for the MAG—R motif ($-2 \log \Lambda = 22.15$, df = 4; $P < 0.005$ in a chi-square distribution). Because the number of observations in the two cells was often less than seven, we simulated nucleotide frequencies in the motif (1,000 trials) under the null hypothesis of random base usage. The null distribution of the likelihood ratio test statistic was then compared with the actual test value of 22.15 (see fig. 4*B*). This analysis confirms the initial result, which showed significant support for the alternative hypothesis of the MAG—R proto-splice site.

To gain a better understanding of the pattern of conservation, chi-square tests were conducted to examine the goodness of fit of the MAG—R hypothesis to the observed intron insertion data. With this analysis, we wanted to determine which of the sites in the MAG—R motif was contributing most strongly to the finding of a significant departure from random base usage. The results of the chi-square analysis showed that the -2 (A), -1 (G), and $+1$ (R) sites differed significantly from the null model, whereas the -3 (M) site showed a base composition that was not significantly different from the null model (fig. 4C). In addition, the $+1$ (R) site in the rRNA proto-splice site is essentially the G nucleotide, and the A contributes little to the signal at this site. Usage of G alone at the $+1$ site in the chi-square analysis resulted in $P < 0.005$ (chi-square value $= 25.98$) in support of the nonrandom model, whereas A alone did not differ significantly from the null model. Our analyses provide, therefore, support for the existence of a proto-splice site in rRNA genes and show that this sequence is closely related to that proposed in pre-mRNA genes.

The close sequence similarity between the Euascomycetes rRNA and pre-mRNA proto-splice sites suggests that our findings are of general importance in understanding spliceosomal intron spread in nuclear coding regions. Interesting questions that still need to be answered concern the extent of the proto-splice site in rRNAs. Inspection of figure 4A shows that in addition to the AG—G motif, positions -8 (G), -7 (G), and $+7$ (A) also show nonrandom nucleotide usage. It is clearly possible that the rRNA proto-splice site extends to bases outside of the proposed AG—G pattern. These more distal positions could play an important role in splice site recognition in pre-RNA that has not yet been recognized. The current data set of 27 distinct intron patterns is, however, not large enough to address this issue. One would, for instance, predict that as one moves farther from the intron insertion site, nucleotide positions will be encountered which show nonrandom usage by chance alone. Also of concern is the fact that the 27 distinct patterns studied here are not drawn from 27 independent genes because of the existence of multiple unique patterns in some genes (e.g., *Gyalecta jenensis*). For these reasons, we restricted the present analysis to testing the a priori hypothesis of a MAG—R proto-splice site. The availability of a larger data set will allow us to ask broader questions about patterns of sequence conservation at spliceosomal intron insertion sites (unpublished data).

Two possible explanations for the conserved sequence flanking rDNA introns are that (1) the AG—G motif is a favored site for intron insertion (proto-splice site model), or (2) the AG—G motif results from strong selection pressure postinsertion, to create a splicing signal required for efficient intron excision (Newman and Norman 1992). To distinguish between these two scenarios, we used as a model the sequences flanking LSU rRNA intron insertion sites in the Euascomycetes with and without introns (32 taxa), as well as eight Ascomycota outside this lineage. This analysis showed that

sequences flanking virtually all of the intron insertion sites (mostly AG—G) were conserved in the outgroup Ascomycota, as well as in Euascomycetes with or without introns. This same pattern was also found with the SSU rRNA introns. An interesting exception was the *G. jenensis* 711 intron, in which 39 LSU rDNA sequences in our data set without an intron at the 711 site had the AG—A motif, whereas Gyalecta, the sole taxon with the 711 intron, had the AG—G motif. This may be an example of postinsertion selection for the AG—G splice site. The only other taxon to have the AG—G motif was *Trapeliopsis granulosa,* which is closely related to Gyalecta (fig. 1).

Taken together, our data are consistent with the idea that most rDNA spliceosomal introns have been inserted (or retained) in conserved regions that primarily encode the AG—G motif. This result supports Dibb and Newman's (1989) hypothesis that proto-splice sites (pre)existed in genes within outgroup taxa lacking introns. Proto-splice sites in rRNA genes appear to mark favored sites of intron insertion rather than sites of intron loss; the 711 site may be an exception (Long et al. 1998). On the basis of this evidence, our working hypothesis is that the fungal rDNA spliceosomal introns have arisen from reverse-splicing and that they have been targeted into a conserved AG—G proto-splice site. An important advantage of our system is that all rDNA introns, whether they are relatively old or new, were presumptively inserted after the origin of the Euascomycetes. Given this, rRNA genes which have never contained introns (i.e., taxa outside the Euascomycetes) can be used to understand the effects of intron insertion on flanking-sequence evolution.

Most of the introns appear to have a recent and restricted distribution within the Euascomycetes, with the exception of the 330 and 393 SSU rRNA and the 678 LSU rRNA introns. It is, however, difficult to determine whether these introns result from multiple "hits" on favored sites or from a single origin in the Euascomycetes ancestor followed by widespread loss. The lack of bootstrap support for nodes joining lineages containing the 330 intron leaves the origin of this intron unresolved. In addition, the presence of introns at positions 330, 331, and 332 suggests that this rRNA region may be prone to multiple independent insertions or intron sliding (e.g., Cubero, Bridge, and Crespo 2000). For these reasons, we did not attempt to account for the origin of each rRNA intron. It is likely that additional intron data and more resolved host trees will ultimately allow us to understand the contributions of rare origins followed by stability, loss, and sliding to the creation of the observed intron distribution.

Intron Positions and rRNA Structure

A prediction of the reverse-splicing model is that rDNA spliceosomal introns should be clustered in regions that are accessible to reverse-splicing "attack" by free introns (e.g., not be hidden by rRNA or ribosome secondary/tertiary structure; Woodson and Cech 1989; Roman and Woodson 1995). These regions should be
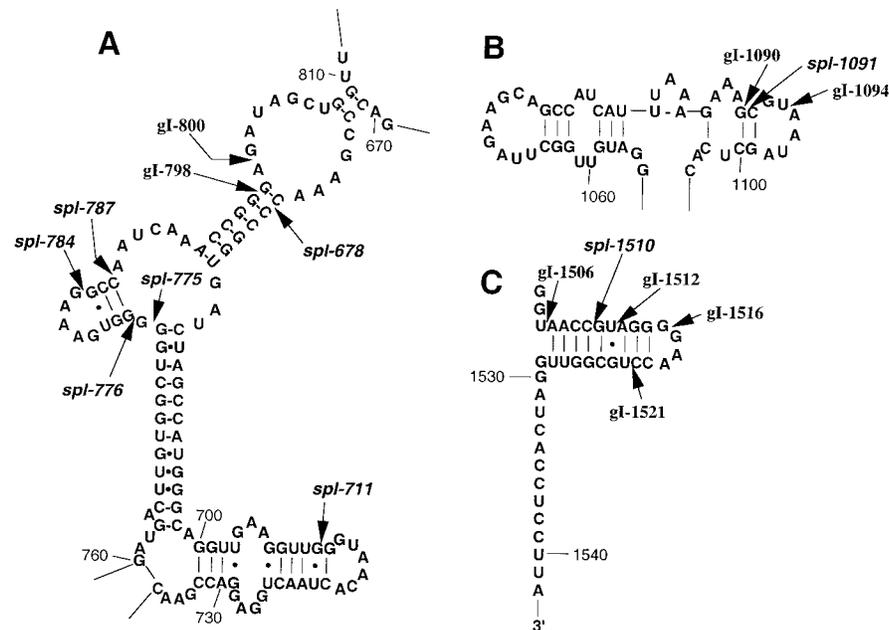
FIG. 5.—Partial secondary structures of the rRNAs of *Escherichia coli*; the positions of spliceosomal (*spl*) and group I (*gI*) introns are indicated with arrows. *A*, LSU rRNA region between 669 and 811. *B*, LSU rRNA region between 1055 and 1104. *C*, SSU rRNA region between 1504 and the 3′ terminus. Note the remarkable clustering of introns in these defined secondary-structure elements.

rich in group I introns which are also believed to spread in nuclear rRNAs through reverse-splicing (Woodson and Cech 1989; Bhattacharya, Friedl, and Damberger 1996). Evidence for reverse-splicing–mediated integration of the *Tetrahymena thermophila* group I intron into preferred sites in *E. coli* rRNAs has been shown in vivo (Roman and Woodson 1998). To test the prediction, we mapped spliceosomal intron sites on *E. coli* rRNA secondary structures (retrieved from the Comparative RNA web site; Gutell 1996). This analysis shows that most of the spliceosomal intron sites are clustered and/or located in close proximity to regions that are interrupted by group I introns, consistent with an origin through reverse-splicing. For example, the 678, 711, 775, 776, 784, and 787 LSU rDNA spliceosomal introns are concentrated in an rRNA secondary-structure region that also harbors two group I introns in ascomycetes (at positions 798 and 800; see fig. 5*A*). Similarly, the 1091 LSU rDNA spliceosomal intron site is bounded by two group I intron sites in the ascomycetes (1090 and 1094; fig. 5*B*; unpublished data), and the 1510 SSU rDNA spliceosomal intron is in close proximity to group I introns at positions 1506, 1512, 1516, and 1521 (fig. 5*C*). In fact, within the LSU rRNA regions that we studied, only three spliceosomal introns (711, 858, and 978) were not found in close proximity (i.e., neighboring sites in primary or secondary structure) to other spliceosomal or group I introns. In the SSU rRNA, only the 265 spliceosomal intron occurred in a region not containing other introns.

Spliceosomal Intron Origin in rDNA Genes

Regarding how the spliceosomal introns may first have appeared in ascomycete rDNA genes, recent anal-

yses show the existence of diffuse, splicing factor–rich nuclear speckles throughout the nucleus (Wei et al. 1999). These data also demonstrate that both speckle-associated and non-speckle-associated regions of the nucleus contain sites for the coordination of transcription and splicing. If we accept that RNA polymerase I–mediated transcription of rRNA genes occurs exclusively in the nucleolus of ascomycetes, then the origin of rDNA spliceosomal introns may be explained by the "intrusion" of splicing factors into this structure. The origin of rDNA introns may have occurred like an "infectious disease" in which the introduction of a single intron, which could reverse-splice into new sites, led to the progressive spread of these sequences into different genic regions. Thereafter, a high concentration of rRNA introns would have increased the nucleolar concentration of splicing factors, thus favoring further spread of introns through reverse-splicing.

Our data suggest that the initial intron invasion likely occurred soon after the origin of the Euascomycetes (fig. 1) and that upon establishment, rDNA introns spread within many lineages. It is important to note that this origin of the spliceosomal intron occurred in a portion of the tree in which the origin(s) of the lichen symbiosis is likely to have taken place (unpublished data). The transition to this symbiotic state was a major event in the evolution of fungi and ascomycetes. One fifth of all fungi form a lichen symbiosis (Hawksworth 1988*a*). More than 98% of the lichen-forming fungi are classified within the Ascomycota, representing about half of the entire Ascomycota diversity (Hawksworth 1988*b*). The Ascomycota is the largest of the four fungal phyla. Except for Chaetosphaeria and Melanochaeta (pyrenomycetes), all species reported to date with spliceosomal

introns have been from lichen-forming fungi or from nonlichenized fungi that are known (unpublished data) to be derived from a lichenized ancestor (e.g., Capronia and Phialophora, Chaetothyriales; see fig. 1).

In conclusion, we suggest that the spliceosome itself and not an external source (e.g., transposable elements, group II introns) may give rise to Euascomycetes rDNA introns through aberrant reverse-splicing of free introns into the abundant rRNA transcripts (as has been suggested for fungal snRNAs; Tani and Ohshima 1989). This is not a new idea (Sharp 1985; Fink 1987; Martinez, Martin, and Cerff 1989; Rzhetsky et al. 1997), but it is one that we are able to extend to a recent intron invasion within taxa of reasonably well known phylogenetic relationships. The close proximity of most rDNA spliceosomal introns to other spliceosomal or group I introns also suggests that particular rRNA regions may be preferred sites for insertion. Our results therefore provide an explanation for the remarkable widespread occurrence of spliceosomal introns in Euascomycetes rDNA genes and implicate reverse-splicing as the mechanism of intron origin. This provides a conceptual basis for addressing, in future studies, both the mechanism and the site of rDNA and, by extension, pre-mRNA spliceosomal intron insertion. An important general implication of our model is that it suggests that the positions of spliceosomal introns in protein-coding genes may also be explained by the combination of a proto-splice site in a pre-mRNA region that is available for reverse-splicing attack by the spliceosome.

Acknowledgments

LITERATURE CITED

BELFORT, M., and P. S. PERLMAN. 1995. Mechanisms of intron mobility. J. Biol. Chem. **270**:30237–30240.

BERGET, S. M. 1995. Exon recognition in vertebrate splicing. J. Biol. Chem. **270**:2411–2414.

BHATTACHARYA, D., T. FRIEDL, and S. DAMBERGER. 1996. Nuclear-encoded rDNA group I introns: origin and evolutionary relationships of insertion site lineages in the green algae/charophytes. Mol. Biol. Evol. **13**:978–989.

BHATTACHARYA, D., S. K. STICKEL, and M. L. SOGIN. 1993. Isolation and molecular phylogenetic analysis of actin coding regions from the prymnesiophyte alga, *Emiliania huxleyi,* using reverse transcriptase and PCR methods. Mol. Biol. Evol. **10**:689–703.

BHATTACHARYA, D., and K. WEBER. 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.

BIDERRE, C., G. METENIER, and C. P. VIVARES. 1998. A small spliceosomal-type intron occurs in a ribosomal protein gene of the microsporidian *Encephalitozoon cuniculi.* Mol. Biochem. Parasitol. **94**:283–286.

CAVALIER-SMITH, T. 1991. Intron phylogeny: a new hypothesis. Trends Genet. **7**:145–148.

CECH, T. R. 1985. Self-splicing RNA: implications for evolution. Int. Rev. Cytol. **93**:3–22.

———. 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. Gene **73**:259–271.

COUSINEAU, B., S. LAWRENCE, D. SMITH, and M. BELFORT. 2000. Retrotransposition of a bacterial group II intron. Nature **404**:1018–1021.

COUSINEAU, B., D. SMITH, S. LAWRENCE-CAVANAGH, J. E. MUELLER, J. YANG, D. MILLS, D. MANIAS, G. DUNNY, A. M. LAMBOWITZ, and M. BELFORT. 1998. Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. Cell **94**:451–462.

CUBERO, O. F., P. D. BRIDGE, and A. CRESPO. 2000. Terminal-sequence conservation identifies spliceosomal introns in ascomycete 18S RNA genes. Mol. Biol. Evol. **17**:751–756.

DE SOUZA, S. J., M. LONG, R. J. KLEIN, S. ROY, S. LIN, and W. GILBERT. 1998. Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proc. Natl. Acad. Sci. USA **95**:5094–5099.

DIBB, N. J., and A. J. NEWMAN. 1989. Evidence that introns arose at proto-splice sites. EMBO J. **8**:2015–2021.

EICKBUSH, T. H. 2000. Molecular biology. Introns gain ground. Nature **404**:940–941.

ESKES, R., J. YANG, A. M. LAMBOWITZ, and P. S. PERLMAN. 1997. Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. Cell **88**:865–874.

FELSENSTEIN, J. 1985. Confidence intervals on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

FINK, G. R. 1987. Pseudogenes in yeast? Cell **49**:5–6.

GARGAS, A., P. T. DEPRIEST, and J. W. TAYLOR. 1995. Positions of multiple insertions in SSU rDNA of lichen-forming fungi. Mol. Biol. Evol. **12**:208–218.

GARGAS, A., and J. W. TAYLOR. 1992. Polymerase chain reaction (PCR) primers for amplifying and sequencing nuclear 18S rDNA from lichenized fungi. Mycologia **84**:589–592.

GILBERT, W. 1978. Why genes in pieces? Nature **271**:501.

GIROUX, M. J., M. CLANCY, J. BAIER, L. INGHAM, D. Mc-CARTY, and L. C. HANNAH. 1994. De novo synthesis of an intron by the maize transposable element *Dissociation.* Proc. Natl. Acad. Sci. USA **91**:12150–12154.

GREEN, R., and H. F. NOLLER. 1997. Ribosomes and translation. Annu. Rev. Biochem. **66**:679–716.

GRUBE, M., A. GARGAS, and P. T. DEPRIEST. 1996. A small insertion in the SSU rDNA of the lichen fungus *Arthonia lapidicola* is a degenerate group-I intron. Curr. Genet. **29**:582–586.

GUTELL, R. R. 1996. Comparative sequence analysis and the structure of 16S and 23S RNA. Pp. 111–128 *in* R. A. ZIMMERMAN and A. E. DAHLBERG, eds. Ribosomal RNA, struc-

ture, evolution, processing and function in protein biosynthesis. CRC Press, Boca Raton, Fla.

HAWKSWORTH, W. L. 1988*a*. The variety of fungal-algal symbioses, their evolutionary significance, and the nature of lichens. Bot. J. Linn. Soc. **96**:3–20.

———. 1988*b*. The fungal partner. Pp. 35–38 *in* M. GALUN, ed. Handbook of lichenology. Vol. 1. CRC Press, Boca Raton, Fla.

HAWKSWORTH, W. L., P. M. KIRK, B. C. SUTTON, and D. N. PEGLER. 1995. Ainsworth & Bisby's dictionary of the fungi. 8th edition. CAB International, Cambridge.

LOGSDON, J. M. JR. 1998. The recent origins of spliceosomal introns revisited. Curr. Opin. Genet. Dev. **8**:637–648.

LOGSDON, J. M. JR., A. STOLTZFUS, and W. F. DOOLITTLE. 1998. Molecular evolution: recent cases of spliceosomal intron gain? Curr. Biol. **8**:R560–R563.

LOGSDON, J. M. JR., M. G. TYSHENKO, C. DIXON, J. D.-JAFARI, V. K. WALKER, and J. D. PALMER. 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: evidence for the introns-late theory. Proc. Natl. Acad. Sci. USA **92**:8507–8511.

LONG, M., S. J. DE SOUZA, C. ROSENBERG, and W. GILBERT. 1998. Relationship between "proto-splice sites" and intron phases: evidence from dicodon analysis. Proc. Natl. Acad. Sci. USA **95**:219–223.

LOPEZ, P. J., and B. SÉRAPHIN. 1999. Genomic-scale quantitative analysis of yeast pre-mRNA splicing: implications for splice-site recognition. RNA **5**:1135–1137.

LUTZONI, F. 1997. Phylogeny of lichen- and non lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. Syst. Biol. **46**:373–406.

LUTZONI, F., P. WAGNER, V. REEB, and S. ZOLLER. 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. Syst. Biol. **49** (in press).

MCCULLOUGH, A. J., and S. M. BERGET. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol. Cell. Biol. **17**:4562–4571.

MARTINEZ, P., W. MARTIN, and R. CERFF. 1989. Structure, evolution and anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J. Mol. Biol. **208**:551–565.

MEDLIN, L., H. J. ELWOOD, S. STICKEL, and M. L. SOGIN. 1988. The characterization of enzymatically amplified 16S-like rRNA coding regions. Gene **71**:491–499.

MIADLIKOWSKA, J., and F. LUTZONI. 2000. Phylogenetic revision of the genus *Peltigera* (lichen-forming Ascomycota) based on morphological, chemical, and large subunit nuclear ribosomal DNA data. Int. J. Plant Sci. **161**:925–958.

MYLLYS, L., M. KÄLLERSJÖ, and A. TEHLER. 1999. Variable sizes of introns in the SSU rDNA in three species of *Roccella* (Arthoniales, Euascomycetes). Curr. Genet. **36**:79–85.

NEWMAN, A. 1997. RNA splicing: out of the loop. Curr. Biol. **7**:R418–R420.

NEWMAN, A. J., and C. NORMAN. 1992. U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. Cell **68**:743–754.

NOUAUD, D., B. BOEDA, L. LEVY, and D. ANXOLABEHERE. 1999. A P element has induced intron formation in *Drosophila.* Mol. Biol. Evol. **16**:1503–1510.

PALMER, J. D., and J. M. LOGSDON JR. 1991. The recent origins of introns. Curr. Opin. Genet. Dev. **1**:470–477.

PATTHY, L. 1996. Exon shuffling and other ways of module exchange. Matrix Biol. **15**:301–310.

PEDERSON, T. 1998. The plurifunctional nucleolus. Nucleic Acids Res. **26**:3871–3876.

PERLMAN, P. S., and M. PODAR. 1996. Reactions catalyzed by group II introns in vitro. Methods Enzymol. **264**:66–86.

PURUGGANAN, M., and S. WESSLER. 1992. The splicing of transposable elements and its role in intron evolution. Genetica **86**:295–303.

REHNER, S. A., and G. J. SAMUELS. 1994. Taxonomy and phylogeny of *Gliocladium* analysed from nuclear large subunit ribosomal DNA sequences. Mycol. Res. **98**:625–634.

ROGERS, S. O., Z. H. YAN, M. SHINOHARA, K. F. LOBUGLIO, and C. J. K. WANG. 1993. Messenger RNA intron in the nuclear 18s ribosomal RNA gene of deuteromycetes. Curr. Genet. **23**:338–342.

ROMAN, J., and S. A. WOODSON. 1995. Reverse splicing of the *Tetrahymena* IVS: evidence for multiple reaction sites in the 23S rRNA. RNA **1**:478–490.

———. 1998. Integration of the *Tetrahymena* group I intron into bacterial rRNA by reverse splicing in vivo. Proc. Natl. Acad. Sci. USA **95**:2134–2139.

ROY, S. W., M. NOSAKA, S. J. DE SOUZA, and W. GILBERT. 1999. Centripetal modules and ancient introns. Gene **238**:85–91.

RYMOND, B. C., C. PIKIELNY, B. SERAPHIN, P. LEGRAIN, and M. ROSBASH. 1990. Measurement and analysis of yeast pre-mRNA sequence contribution to splicing efficiency. Methods Enzymol. **181**:122–147.

RZHETSKY, A., F. J. AYALA, L. C. HSU, C. CHANG, and A. YOSHIDA. 1997. Exon/intron structure of aldehyde dehydrogenase genes supports the "introns-late" theory. Proc. Natl. Acad. Sci. USA **94**:6820–6825.

SALDANHA, R., G. MOHR, M. BELFORT, and A. M. LAMBOWITZ. 1993. Group I and group II introns. FASEB J. **7**:15–24.

SHARP, P. A. 1985. On the origin of RNA splicing and introns. Cell **42**:397–400.

SOKAL, R. E., and F. J. ROHLF. 1995. Biometry. 3rd edition. W. H. Freeman, New York.

SPATAFORA, J. W., T. G. MITCHELL, and R. VILGALYS. 1995. Analysis of genes coding for small subunit rRNA sequences in studying phylogenetics of dematiaceous fungal pathogens. J. Clin. Microbiol. **33**:1322–1326.

STENROOS, S., and P. T. DEPRIEST. 1998. Small insertions at a shared position in the SSU rDNA of lecanorales (lichen-forming ascomycetes). Curr. Genet. **33**:124–130.

STEPHENS, R. M., and T. D. SCHNEIDER. 1992. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. J. Mol. Biol. **228**:1124–1136.

SWOFFORD, D. L. 1999. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4.0b2. Sinauer, Sunderland, Mass.

TAKAHASHI, Y., S. URUSHIYAMA, T. TANI, and Y. OHSHIMA. 1993. An mRNA-type intron is present in the *Rhodotorula hasegawae* U2 small nuclear RNA gene. Mol. Cell. Biol. **13**:5613–5619.

TANI, T., and Y. OHSHIMA. 1989. The gene for the U6 small nuclear RNA in fission yeast has an intron. Nature **337**:87–90.

———. 1991. mRNA-type introns in U6 small nuclear RNA genes: implications for the catalysis in pre-mRNA splicing. Genes Dev. **5**:1022–1031.

TOMITA, M., N. SHIMUZU, and S. BRUTLAG. 1996. Introns and reading frames: correlation between splicing sites and their codon positions. Mol. Biol. Evol. **13**:11–15.

VAN ETTEN, J. L., and R. H. MEINTS. 1999. Giant viruses infecting algae. Annu. Rev. Microbiol. **53**:447–494.

VILGALYS, R., and M. HESTER. 1990. Rapid genetic identification and mapping enzymatically amplified ribosomal DNA from several *Cryptococcus* species. J. Bacteriol. **172**: 4238–4246.

WEBER, K., and W. KABSCH. 1994. Intron positions in actin genes seem unrelated to the secondary structure of the protein. EMBO J. **13**:1280–1286.

WEI, X., S. SOMANATHAN, J. SAMARABANDU, and R. BEREZNEY. 1999. Three-dimensional visualization of transcription sites and their association with splicing factor-rich nuclear speckles. J. Cell Biol. **146**:543–548.

WINKA, K., C. AHLBERG, and O. E. ERIKSSON. 1998. Are there lichenized Ostropales? J. Lichenol. (Lond.) **30**:455–462.

WOODSON, S. A., and T. R. CECH. 1989. Reverse self-splicing of the *Tetrahymena* group I intron: implication for the directionality of splicing and for intron transposition. Cell **57**: 335–345.

ZIMMERLY, S., H. GUO, R. ESKES, J. YANG, P. S. PERLMAN, and A. M. LAMBOWITZ. 1995. A group II intron RNA is a catalytic component of a DNA endonuclease involved in intron mobility. Cell **83**:529–538.

ZOLLER, S., F. LUTZONI, and C. SCHEIDEGGER. 1999. Genetic variation within and among populations of the threatened lichen *Lobaria pulmonaria* in Switzerland and implications for its conservation. Mol. Ecol. **8**:2049–2059.