RNA (2002), *8*:1373–1377. Cambridge University Press. Printed in the USA.

DOI: 10.1017/S1355838202023087

LETTER TO THE EDITOR

Identification of a family of group II introns encoding LAGLIDADG ORFs typical of group I introns

NAVTEJ TOOR and STEVEN ZIMMERLY
Department of Biological Sciences, University of Calgary, Calgary, Alberta T2N 1N4, Canada

ABSTRACT

Group I and group II introns are unrelated classes of introns that each encode proteins that facilitate intron splicing and intron mobility. Here we describe a new subfamily of nine introns in fungi that are group II introns but encode LAGLIDADG ORFs typical of group I introns. The introns have fairly standard group IIB1 RNA structures and are inserted into three different sites in SSU and LSU rRNA genes. Therefore, introns should not be assumed to be group I introns based solely on the presence of a LAGLIDADG ORF.

Keywords: catalytic RNA; fungi; intron-encoded protein; mitochondria; ribozyme

INTRODUCTION

Despite being unrelated in sequence and structure, group I and group II introns share many properties (Lambowitz et al., 1999; Bonen & Vogel, 2001; Belfort et al., 2002). Both types of introns are found in genomes of lower eukaryotic organelles and in bacteria, both are capable of self-splicing in vitro, and both encode proteins that aid in splicing of the introns in vivo and sometimes confer mobility to the introns. One difference between group I and group II introns is that group I introns encode at least four classes of ORFs, all nucleases, which appear to have invaded the introns many times (Gimble, 2000; Chevalier & Stoddard, 2001; Belfort et al., 2002), whereas group II introns are associated with a single ORF type, a reverse transcriptase (RT), which appears to have primarily coevolved with the intron RNA structure (Toor et al., 2001). Here we describe an exception in which group II introns encode LAGLIDADG ORFs typical of group I introns rather than RT-related proteins.

RESULTS AND DISCUSSION

During our analysis of ORFs encoded within group II introns, we discovered a group II intron encoding an ORF of the LAGLIDADG family (*Cryphonectria parasitica* SSUI1, Table 1). Folding of the intron confirmed it to have a group IIB1 structure (Fig. 1A), and the ORF contained two LAGLIDADG motifs typical of group I intron-encoded proteins. Previously, it had been noted that a LAGLIDADG ORF was present in a partial group II intron sequence from *Trimorphomyces papilionaceus* (Michel & Ferat, 1995), and so we undertook a broader search to determine the number of introns of this type.

Relatives of the *C. parasitica* SSUI1 were found by searching GenBank with BLASTN, using queries of the intron RNA sequence (omitting the ORF). The resulting matches were then screened for the presence of an ORF of any type using BLASTX. Iterative cycles of searches, using newly identified introns as queries, identified a total of nine group II introns that encode LAGLIDADG-related ORFs (Table 1). Five of the introns are closely related homologs in *Cordyceps* species. In addition, two other introns were found, *Grifola frondosa* SSUI1 and *Lentaria byssiseda* SSUI1, which appear to be homologs of LAGLIDADG-encoding *C. parasitica* SSUI1; however, only C-terminal ORF remnants remain without the LAGLIDADG motifs (Table 1 footnote).

Because the search strategy was based on finding relatives of a starting intron, a more general search was done. Plant organelles were screened for this type of intron by searching plant genomes in GenBank by TBLASTN using queries of LAGLIDADG ORFs from Table 1, but only known LAGLIDADG-encoding group I

Reprint requests to: Dr. Steven Zimmerly, Department of Biological Sciences, University of Calgary, 2500 University Drive NW, Calgary, Alberta T2N 1N4, Canada; e-mail: zimmerly@acs.ucalgary.ca.

TABLE 1. Group II introns that encode LAGLIDADG ORFs[a]

Species	Phylum[b]	Host gene	Intron No.	Intron RNA domains	ORF size (amino acids)	ORF frameshifts and stop codons[c]	Accession No.
Agrocybe aegerita	B	LSU rRNA	5	1–6	320	3 frameshifts, 3 stop codons	AF087656
Cryphonectria parasitica	A	SSU rRNA	1	1–6	286	none	AF029891
Cryphonectria parasitica	A	SSU rRNA	3	1–6	353	none	AF029891
Cordyceps konnoana	A	SSU rRNA	1	1–6	344	none	AB031194
Cordyceps ramosopulvinata	A	SSU rRNA	1	1–6	313	none	AB027348
Cordyceps sobolifera	A	SSU rRNA	1	1–6	311	1 stop codon	AB027350
Cordyceps sp. 97003	A	SSU rRNA	1	1–6	376	4 frameshifts, 2 stop codons	AB027352
Cordyceps sp. 97009	A	SSU rRNA	1	1–6	310	none	AB027356
Trimorphomyces papilionaceus	B	SSU rRNA	1	5,6 (T)[d]	263	none	X73671

[a]Two introns are omitted from the table because their ORFs are degenerate fragments that lack LAGLIDADG motifs. *Lentaria byssiseda* SSUI1 (AF185983) and *Grifola frondosa* SSUI1 (AF334880) are both putative homologs of *C. parasitica* SSUI1. Full group II intron structures are present, but the *L. byssiseda* ORF fragment encodes only 88 amino acids with similarity to the *C. parasitica* SSUI1 ORF (with one premature stop codon), while *G. frondosa* SSUI1 encodes 104 amino acids with similarity to the *C. parasitica* SSUI1 ORF (with three frame shifts). *G. frondosa* and *L. byssiseda* are both basidiomycetes.
[b]A: Ascomycota, B: Basidiomycota.
[c]The ORFs were translated using NCBI ORF Finder (genetic code #4).
[d](T) indicates a truncation due to a partial sequence.

introns were found. To identify LAGLIDADG-encoding group II introns in fungi, protists, and other nonplants, all organellar introns (both group I and group II introns) were downloaded from The Organellar Genome Database (GOBASE; Shimko et al., 2001) along with 20 bp of flanking sequences. The intron data set was screened for group II intron domain 5 motifs using the program RNAMotif (Macke et al., 2001). The descriptor for domain 5 consisted of a 9-bp stem beginning with 5′NNGC, a second stem of 3–5 bp with a 4–5 nt loop, and a 2-nt bulge of any sequence in place of the AC bulge. The resulting group II introns were then searched for LAGLIDADG ORFs using BLASTX, which yielded the introns *C. parasitica* SSUI1, *C. parasitica* SSUI3, *Agrocybe aegerita* LSUI5 and *Trimorphomyces papilionaceus* SSUI1. The remaining introns in Table 1 are not represented in GOBASE. Finally, bacterial sequences in GenBank were screened with BLASTN for intron domain 5, or with BLASTX for LAGLIDADG ORFs, using a variety of intron and LAGLIDADG queries, but only known group II introns were identified, none encoding LAGLIDADG ORFs. We conclude that the introns in Table 1 are a fairly complete listing of introns of this type that are currently sequenced.

Correct boundaries of the introns were confirmed for all of the introns by comparing flanking rRNA sequences with related rRNA sequences lacking introns; however, for *A. aegerita* LSUI5, although the 3′ boundary is clear, the 5′ boundary is ambiguous. We placed the 5′ terminus 20 bp downstream of the predicted break point in rRNA alignments, which allowed folding into a standard group II intron structure (Fig. 1C), although the structure lacks strong IBS1/EBS1 and IBS2/EBS2 pairings.

None of the introns are properly annotated in GenBank. The *Cordyceps* introns are annotated as portions of SSU rRNA genes without identification of introns or ORFs; *C. parasitica* SSUI1 and *C. parasitica* SSUI3 are annotated as introns encoding LAGLIDADG ORFs but with incorrect intron boundaries, perhaps according to group I intron structures; *A. aegerita* LSUI5 is annotated as an “intron” with incorrect boundaries and without ORF annotation.

The introns folded into fairly standard group IIB1 intron structures, with the ORFs located in domain IV (Fig. 1). Irregularities include the absence of good IBS1/EBS1 and IBS2/EBS2 pairings in *A. aegerita* LSUI5, a very weak IBS1/EBS1 pairing and no discernible IBS2/EBS2 pairing for *C. parasitica* SSUI1, and weak IBS2/EBS2 pairings for *Cordyceps* introns and *C. parasitica* SSUI3. There is little sequence conservation that would indicate a close relationship between the intron structures, except for the five *Cordyceps* introns and *C. parasitica* SSUI3 intron, which are >75% identical.

Phylogenetic analysis was performed on the nine essentially complete ORFs after correcting for frameshifts and stop codons, using neighbor-joining and maximum parsimony algorithms. Although substantial ORF degeneration limits reliability of the inferred relationships, nevertheless, three phylogenetic groupings are suggested which correspond to three different insertion sites (Fig. 2).

The LAGLIDADG ORFs are not closely related to any other LAGLIDADG ORFs, and so their origins are uncertain. Their closest relatives, judging from matches in BLAST searches, appear to be diverse fungal LAGLIDADG ORFs. This is consistent with the possibility that the ORFs invaded group II introns in fungal mitochondria. It is interesting that the ORFs are all located in domain IV, as are RT-related ORFs in group II introns. This may be due to functional constraints that prevent ORF insertions elsewhere in the structure, or may be due to the lengthy spacers in domain IV of many ORF-less group II introns (Toor et al., 2001), which would provide a large target for ORF invasion.

https://doi.org/10.1017/S1355838202023087 Published online by Cambridge University Press

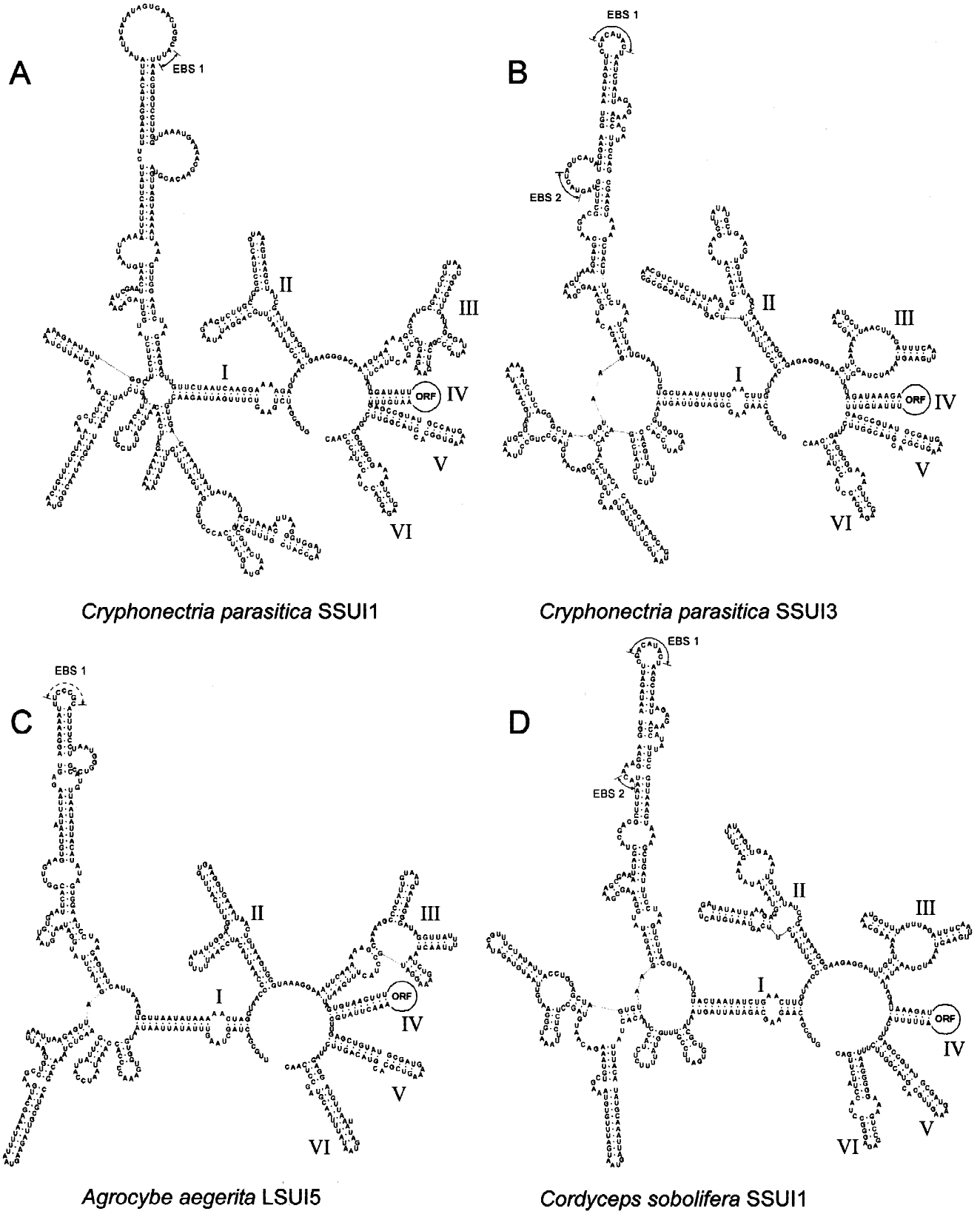

FIGURE 1. Secondary structure models of group II introns encoding LAGLIDADG ORFs. All structures are fairly standard IIB1 structures, with exceptions noted in the text. Only one *Cordyceps* intron structure is shown because the five structures are nearly identical. The positions of EBS1 and EBS2 are noted, with the *A. aegerita* EBS1 dotted because it does not pair well with IBS1. All RNA structures were folded using Mfold version 3.1 (Mathews et al., 1999; Zuker et al., 1999) with manual refinement using RNAdraw (Matzura & Wennborg, 1996) and final manipulation with RNAViz (De Rijk & De Wachter, 1997).

https://doi.org/10.1017/S1355838202023087 Published online by Cambridge University Press

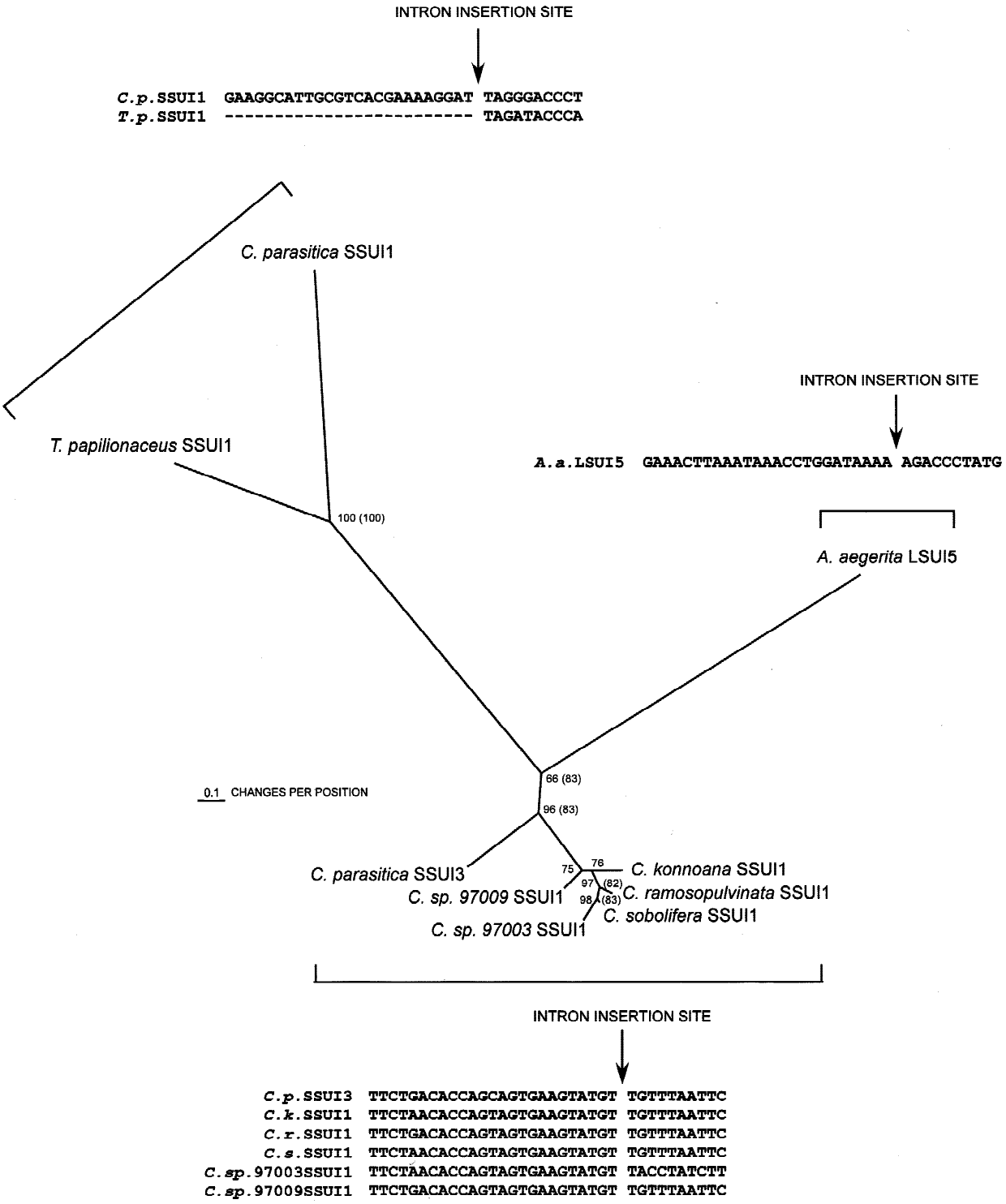

FIGURE 2. Phylogenetic groupings of intron-encoded ORFs. Three suggested ORF groupings coincide with three intron insertion sites as indicated. For phylogenetic analysis, ORFs were first corrected for frameshifts and premature stop codons, and then aligned by CLUSTALX with manual refinement. Regions with poor alignment were omitted, giving a total data set of 250 amino acids. Phylogenetic relationships were inferred by neighbor-joining and maximum parsimony algorithms of PHYLIP version 3.573c (Felsenstein, 1985, 1995). Neighbor-joining analysis was with the programs SEQBOOT, PROTDIST, NEIGHBOR, and CONSENSE. Maximum parsimony analysis utilized the programs SEQBOOT, PROTPARS, and CONSENSE. The tree shown is the single best tree from neighbor-joining without bootstrapping, and bootstrap values are indicated for the majority rule consensus tree for neighbor-joining (normal type) and maximum parsimony analyses (parentheses).

https://doi.org/10.1017/S1355838202023087 Published online by Cambridge University Press

An intriguing unanswered question is whether a LAGLIDADG ORF invaded group II introns only once, followed by divergence into three homing sites, or whether the invasion occurred three times into three different intron structures. To address this issue, we phylogenetically analyzed LAGLIDADG ORFs present in the Pfam database (Bateman et al., 2002); however, the data did not distinguish whether the three groupings of ORFs in Figure 2 are more closely related to each other than to other LAGLIDADG ORFs (not shown). On one hand, a single ORF invasion followed by divergence might seem likely because all introns are IIB1 RNA structures and the ORF is located in domain IV in all cases. On the other hand, the very frequent ORF degeneration (5 out of 11 examples in Table 1, including related *Cordyceps* introns), argues against a stable long-term association between introns and ORFs. The frequency of ORF degeneration also casts doubt on maturase activities of the LAGLIDADG proteins, which would be expected to result in maintenance of the ORF. Similarly, it is not clear whether endonuclease activities of the ORFs would provide mobility to these introns.

LAGLIDADG ORFs are known to be promiscuous ORFs that insert into a variety of elements, including group I introns, archaebacterial introns, and inteins, and they can also exist as free-standing ORFs (Gimble, 2000; Chevalier & Stoddard, 2001; Belfort et al., 2002). It has been rationalized that the ORFs are independent mobile elements that insert into the "safe havens" of introns or inteins (Loizos et al., 1994), and in this vein of thought, the LAGLIDADG invasion of group II introns might simply reflect the invasive properties of the ORFs, and the ORFs may not necessarily contribute to intron splicing or mobility.

Finally, because none of these group II introns was correctly identified, it would be wise in the future not to assume introns are group I introns solely on the basis of LAGLIDADG ORFs. Care should be taken especially in fungal genomes, where all known basidiomycete group II introns are of this type. In distinguishing between group I and group II introns, domain 5 of group II introns is highly conserved in sequence and structure and is the best indicator of the presence of a group II intron.

NOTE ADDED IN REVISION

Since our analysis, another similar intron has been reported in *Rhizophydium sp.* mitochondria (GenBank accession number AF404306). The intron is located in the *cox1* gene rather than a rRNA gene.

NOTE ADDED IN PROOF

Georg Mohr has brought to our attention that the secondary structures of the five *Cordyceps* introns and *C.p.*SSUI3 can be drawn in an alternate arrangement with the ORF located in a domain III. This alternate structure is supported by comparison between the *C.p.*SSUI3 and *Cordyceps* sequences, and is more consistent with the expected domain III structure. If the revised structure is correct, this would be the first example of a group II intron ORF located outside of domain IV, and it would also indicate that the LAGLIDADG ORFs inserted independently into these introns compared to *C.p.*SSUI1 and *A.a.*LSUI5, whose ORFs are clearly located in domain IV.

Received July 2, 2002; returned for revision July 16, 2002; revised manuscript received August 14, 2002

REFERENCES

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. 2002. The Pfam protein families database. *Nucleic Acids Res 30*:276–280.

Belfort M, Derbyshire V, Parker MM, Cousineau B, Lambowitz, AM. 2002. Mobile introns: Pathways and proteins. In: Craig NL, Craigie R, Gellert M, Lambowitz AM, eds. *Mobile DNA II*. Washington, DC: ASM Press. pp 761–783.

Bonen L, Vogel J. 2001. The ins and outs of group II introns. *Trends Genet 6*:322–331.

Chevalier BS, Stoddard BL. 2001. Homing endonucleases: Structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res 29*:3757–3774.

De Rijk P, De Wachter R. 1997. RnaViz, a program for the visualisation of RNA secondary structure. *Nucleic Acids Res 25*:4679–4684.

Felsenstein J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution 39*:783–791.

Felsenstein J. 1995. PHYLIP (Phylogeny Inference Package) version 3.55c. Distributed by the author. Department of Genetics, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip/getme.html.

Gimble FS. 2000. Invasion of a multitude of genetic niches by mobile endonuclease genes. *FEMS Microbiol Lett 185*:99–107.

Lambowitz AM, Caprara M, Zimmerly S, Perlman PS. 1999. Group I and group II ribozymes as RNPs: Clues to the past and guides to the future. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world*, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 451–485.

Loizos N, Tillier ER, Belfort M. 1994. Evolution of mobile group I introns: Recognition of intron sequences by an intron-encoded endonuclease. *Proc Natl Acad Sci USA 91*:11983–11987.

Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Res 29*:4724–4735.

Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol 288*:911–940.

Matzura O, Wennborg A. 1996. RNAdraw: An integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *CABIOS 12*:247–249.

Michel F, Ferat JL. 1995. Structure and activities of group II introns. *Annu Rev Biochem 64*:435–461.

Shimko N, Liu L, Lang BF, Burger G. 2001. GOBASE: The organelle genome database. *Nucleic Acids Res 29*:128–132.

Toor N, Hausner G, Zimmerly S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA 7*:1142–1152.

Zuker M, Mathews DH, Turner DH. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In: Barciszewski J, Clark BFC, eds. *RNA biochemistry and biotechnology*. Dordrecht, The Netherlands: Kluwer Academic Publishers. pp 11–43.