

# The Evolution of Homing Endonuclease Genes and Group I Introns in Nuclear rDNA

Peik Haugen,\* Valérie Reeb,† François Lutzoni,† and Debashish Bhattacharya\*

\*Department of Biological Sciences and Center for Comparative Genomics, University of Iowa; and

†Department of Biology, Duke University

Group I introns are autonomous genetic elements that can catalyze their own excision from pre-RNA. Understanding how group I introns move in nuclear ribosomal (r)DNA remains an important question in evolutionary biology. Two models are invoked to explain group I intron movement. The first is termed homing and results from the action of an intron-encoded homing endonuclease that recognizes and cleaves an intronless allele at or near the intron insertion site. Alternatively, introns can be inserted into RNA through reverse splicing. Here, we present the sequences of two large group I introns from fungal nuclear rDNA, which both encode putative full-length homing endonuclease genes (HEGs). Five remnant HEGs in different fungal species are also reported. This brings the total number of known nuclear HEGs from 15 to 22. We determined the phylogeny of all known nuclear HEGs and their associated introns. We found evidence for intron-independent HEG invasion into both homologous and heterologous introns in often distantly related lineages, as well as the “switching” of HEGs between different intron peripheral loops and between sense and antisense strands of intron DNA. These results suggest that nuclear HEGs are frequently mobilized. HEG invasion appears, however, to be limited to existing introns in the same or neighboring sites. To study the intron-HEG relationship in more detail, the S943 group I intron in fungal small-subunit rDNA was used as a model system. The S943 HEG is shown to be widely distributed as functional, inactivated, or remnant ORFs in S943 introns.

## Introduction

Group I introns are autonomous genetic elements that have a characteristic RNA fold consisting generally of 10 paired elements (P1 to P10). These conserved RNA regions catalyze a two-step, self-splicing reaction resulting in intron release and ligation of the exons (Cech 1990). When they self-splice, group I introns presumably confer no phenotype, regardless of their insertion site, and may be “silent” parasites able to maintain themselves or spread in genomes. If they lose self-splicing ability, group I introns (i.e., the organisms containing the introns) are likely eliminated because of strong selection against nonfunctional gene products (Dujon 1989).

The nuclear genomes of protists and fungi are rich sources of group I introns and offer an excellent model for understanding group I intron, and generally, autocatalytic RNA evolution (Bhattacharya 1998). Nuclear group I introns are restricted to the ribosomal (r)RNA genes at 170 insertion sites that are approximately equally distributed among the small-subunit (SSU) and large-subunit (LSU) rRNA coding regions. The growing list of these introns has been organized into a Web-based database that currently contains about 1,200 intron sequences from nuclear rDNAs (Cannone et al. 2002; <http://www.rna.icmb.utexas.edu>). Based on secondary structural differences in the peripheral regions, the majority of nuclear group I introns have been categorized into one of two subgroups called group IC1 and group IE introns, where IC1 is the most common in nature. The broad and sporadic distribution of these introns suggests that they have successfully spread into different genomes and genic sites and that they are prone to loss over evolutionary time (Lambowitz and

Belfort 1993; Goddard and Burt 1999; Bhattacharya, Cannone, and Gutell 2001; Bhattacharya, Friedl, and Helms 2002).

Two mechanisms are currently invoked to explain group I intron mobility. The first is homing and is initiated by an intron-encoded endonuclease (ENase) that recognizes and cleaves an intronless allele at or near the intron insertion site (reviewed in Chevalier and Stoddard [2001]). After ENase cleavage at a specific 15-nt to 45-nt target sequence, the intron-containing allele is used as the template in a double-strand break and repair (DSBR) pathway resulting in insertion of the intron and coconversion of flanking exon sequences (Dujon 1989; Belfort and Perlman 1995). Using the homing endonuclease gene (HEG)-associated  $\omega$  intron from the yeast mitochondrion, Goddard and Burt (1999) postulated that group I introns are recurrently gained, degenerate, and are lost in a cyclic manner. Although the intron with HEG is destined to be eventually lost from a population after all individuals are fixed for these elements, the intact intron HEG can remain active if it “escapes” to an intronless population of the same or closely related species to restart the homing cycle.

In contrast to homing that proceeds via a DNA intermediate, ribozyme-mediated intron transfer (referred to as “reverse splicing”) is believed to promote intron mobility through an RNA intermediate (Woodson and Cech 1989; Roman and Woodson 1995; Roman and Woodson 1998; Roman, Rubin, and Woodson 1999). Group I ribozymes recognize their target sequence through complementary base pairing with a short (4 to 6 nt) internal guide sequence. Potentially, the reverse splicing pathway provides a greater possibility of intron movement into heterologous sites in comparison with the high sequence specificity required for homing. This hypothesis has been supported by biochemical and phylogenetic data (Roman, Rubin, and Woodson 1999; Bhattacharya, Friedl, and Helms 2002). Unlike HEGs, it is however unlikely that reverse splicing would be efficient enough to

Key words: Group I introns, homing endonucleases, HEGs, intron mobility, ribosomal RNA.

E-mail: dbhatac@blue.weeg.uiowa.edu.

*Mol. Biol. Evol.* 21(1):129–140, 2004

DOI: 10.1093/molbev/msh005

*Molecular Biology and Evolution* vol. 21 no. 1

© Society for Molecular Biology and Evolution 2004; all rights reserved.

**Table 1**  
**Distribution of Group I Introns with HEGs in Nuclear rDNA**

Organism	Introns <sup>a</sup>	HEGs <sup>b</sup>	rDNA Position <sup>c</sup>
Fungi	847	13	S943(7), S1199(2), S1506(3), L1923
Rhodophyta	155	2	S516, S1506
Chlorophyta	140	1	S943
Acanthamoebidae	14	1	S516
Heterolobosea	29	3	S516, L1926, L2563
Mycetozoa	13	2	S956, L1925
Other	74	0	
Total	1,272	22	

<sup>a</sup> Total number of group I introns in these taxa.

<sup>b</sup> Total number of HEG-containing group I introns.

<sup>c</sup> Intron position in the rDNA gene (S and L are for the small and large subunit rDNA, respectively). The number of HEGs available in these introns (if > 1) is also shown.

maintain introns at high frequencies in populations because of its reliance on chance integration, reverse transcription, and recombination to promote spread. As an additional layer of complexity, both sequence and biochemical data support the idea that HEGs are themselves mobile genetic elements (e.g., Mota and Collins 1988; Lambowitz and Belfort 1993; Loizos, Tillier, and Belfort 1994; Ogawa et al. 1997; Pellenz et al. 2002). By invading group I introns, HEGs ensure rapid propagation of both the group I intron and the HEG. Invasion of HEGs into noncoding, autocatalytic introns also ensures that integration does not inactivate an important host coding region.

Homing ENases encoded by HEGs are divided into four families (GIY-YIG, LAGLIDADG, HNH, and His-Cys box) based on conserved protein motifs (for review, see Chevalier and Stoddard [2001]). The His-Cys box family is exclusively associated with nuclear group I introns. However, only about 22 different His-Cys box ORFs have thus far been identified (Haugen, De Jonckheere, and Johansen 2002; Tanabe, Yokota, and Sugiyama 2002; Yokoyama, Yamagishi, and Hara 2002; this work) in contrast to about 1,200 nuclear group I introns. The distribution of nuclear group I introns and their associated HEGs is shown in table 1.

Here we present the sequences of two novel and five previously unrecognized nuclear HEGs, thereby raising the total number of nuclear HEGs from 15 to 22 (similar HEGs from closely related species are counted as one). The small-subunit (S)943 rDNA group I intron in fungi contains HEGs in seven different species, making this a model for understanding intron-HEG evolution. We have reconstructed the phylogeny of the fungal hosts, their group I introns, and HEGs, with particular emphasis on the S943 introns, to address the following questions: (1) Does the S943 HEG distribution primarily favor vertical ancestry of the ORF, or are these sequences frequently laterally transferred between introns in the same or different species? (2) Are the S943 introns frequently laterally transferred among fungi? (3) How frequently are HEGs transferred to ectopic sites, and if so, in which sites are they fixed in nature?

## Materials and Methods

### Fungal Cultures, DNA Extraction, and PCR Amplification

The two HEGs included in this study, I-*PchI* and I-*CpiI*, were isolated from a herbarium specimen of *Pleopsidium chlorophanum* (Reeb, VR 13-VIII-98/5, DUKE) and a culture of *Capronia pilosella* (S. Huhndorf, F. Fernandez, and M. Huhndorf, SMH2565, The Field Museum, Chicago), respectively. DNA was isolated using the Puregene Kit (Gentra Systems) following the manufacturer's protocol for filamentous fungi. The nuclear S943 group I intron from *Pleopsidium chlorophanum* and *Capronia pilosella* rDNA was amplified by PCR (polymerase chain reaction) in 50 µl standard reactions using 1 µl of genomic DNA template and the fungal-specific nssu1088R and NS22 (Kauff and Lutzoni 2002; Gargas and Taylor 1992) primers. PCR products were purified from agarose gels using the GELase enzyme (Epicentre technologies) and directly sequenced on both strands by using the ABI PRISM 3700 DNA Analyzer (Applied Biosystems) and the Big Dye terminator chemistry (Applied Biosystems).

### Novel HEGs Identified in GenBank and Retrieval of His-Cys Box Protein Sequences

Most of the HEG protein sequences were retrieved from the National Center for Biotechnology Information (NCBI) GenBank (Benson et al. 2003) using published work (e.g., Haugen, De Jonckheere, and Johansen 2002; Tanabe, Yokota, and Sugiyama 2002; Yokoyama, Yamagishi, and Hara 2002) as a guide. However, inspection of the Comparative RNA Web site (<http://www.rna.icmb.utexas.edu/>) revealed several unexpectedly large fungal introns at position S1506. The longest potential protein sequence found in the S1506 intron of *Arthrobotrya superba* (Orbiliomycetes; accession number U51949) was used as the query in a TblastN search at the NCBI Web site (<http://www.ncbi.nlm.nih.gov/BLAST/>). The query sequence showed high identity to potential proteins encoded by the S1506 intron in the fungi *Protomyces pachydermus* (Taphrinomycetes; accession number D85142) and *Tilletiopsis oryzaicola* (Basidiomycota; accession number AB045705). Additional analyses revealed that all three introns contained similar HEG pseudogenes. The *A. superba* and *P. pachydermus* HEGs are located on the sense strand, whereas the *T. oryzaicola* intron is found on the antisense strand. The fungal His-Cys box proteins of *Pseudohalonestria lignicola* (Sordariomycetes; accession number U31812), and *Cordyceps pseudomilitaris* (Sordariomycetes; accession number AF327394) were also found by manual inspection of the Comparative RNA Web site. Finally, representatives from all the major lineages of His-Cys box proteins were used as queries in TblastN searches to identify as many nuclear HEGs as possible. The HEGs are however highly divergent sequences, and because many of them are remnants/pseudogenes, we cannot exclude the possibility that there still remain nuclear HEGs to be discovered in the databases.

The GenBank accession numbers for the *Pleopsidium chlorophanum* and *Capronia pilosella* group I intron sequences are AY316151 and AY316152, respectively.

#### Sequence Alignment and Phylogenetic Analysis of HEGs

His-Cys box protein sequences were manually aligned using the BioEdit version 5.0.9 (Hall 1999) program. Known residues/motifs important for zinc coordination and DNA binding in the I-*PpoI* structure were used as landmarks in the alignment (Flick et al. 1998). The His-Cys box protein data set was, however, very challenging for two reasons. First, the sequences between the conserved residues/domains were highly divergent, making it difficult in some cases to recognize peripheral zinc coordinating residues. Second, about 50% of the HEGs identified in this work were inferred to be nonfunctional because of premature stop codons, the absence of an ATG start codon, frameshift mutations, deletions, or a combination of these changes. When possible, the protein sequences were “repaired” in silico by manually introducing frameshifts or by ignoring stop codons, as previously described for the His-Cys box pseudogenes in red algae (Haugen et al. 1999; Müller et al. 2001). Changes were made on the basis of comparisons of all frames of the protein in question with the sequence of other, preferentially evolutionarily closely related and functional His-Cys box homing ENases. Unalignable sequences, mostly outside of the regions corresponding to the DNA-binding domain and the regions important for zinc coordination in the I-*PpoI* crystal structure (Flick et al. 1998), were excluded from the analyses. The final data set used in the phylogenetic analyses contained 119 aa spanning approximately the first to the last zinc-coordinating residues (Cys 41 and 138 in I-*PpoI*), except the excluded regions. The HEG alignment with all changes annotated is available as Supplementary Material online from the MBE Web site.

Four different methods were used to reconstruct the phylogeny of the divergent His-Cys box endonucleases. First, we used the protein maximum-likelihood (proml) method in PHYLIP version 3.6a3 (Felsenstein 2002) with the JTT +  $\Gamma$  ( $\alpha = 0.45$  [estimated with MrBayes; see below]) evolutionary model (Jones, Taylor, and Thornton 1992) to infer a tree. Five rounds of random addition and global optimization were used to search for the best phylogeny. Support for nodes in the proml tree was assessed with the bootstrap method (200 replications [Felsenstein 1985]) using PHYLIP as described above except that a single round of random addition was used for each bootstrap data set. Second, we used the MEGA version 2.1 program (Kumar et al. 2001) to infer a minimum-evolution tree with Poisson correction as the distance model. Missing sites and gaps, when present, were excluded for each pair during the distance matrix calculation. The maximum number of trees to retain was set to 10, with the close neighbor interchange at search level 2, as the search method. An initial tree was obtained by neighbor-joining. The same minimum evolution settings were used to test the inferred tree in a bootstrap

analysis using 1,000 replicates. Third, we did Bayesian analysis of the HEG data (MrBayes version 3 [Huelsenbeck and Ronquist 2001]) using the WAG + G model (Whelan and Goldman 2001). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) from a random starting tree was initiated in the Bayesian inference and run for 2,000,000 generations with trees sampled every 100th generation. After discarding the first 10,000 trees, a consensus tree was made with the remaining 10,000 phylogenies sampled to determine the posterior probabilities at the different nodes. The average value of the alpha parameter from the 10,000 sampled trees ( $0.45 \pm 0.05$  standard deviation) was used in the proml analysis. Fourth, PAUP\* version 4.0b8 (Swofford 2002) was used to perform unweighted maximum-parsimony bootstrap (1,000 replications) analysis with the HEG data. Heuristic searches were done with each bootstrap data set using the tree bisection-reconnection (TBR) branch-swapping algorithm to find the shortest trees. The number of random-addition replicates was set to 10 for each tree search. The S1199 ENase sequences were chosen arbitrarily to root the HEG tree.

#### Sequence Alignment and Phylogenetic Analysis of SSU rDNA 943 Group I Introns

The fungal (except the intron in the Ericoid mycorrhizal species) S943 group I introns (shown in table 2 of Supplementary Material online) were manually aligned using BioEdit. The alignment was created through juxtaposition of the secondary structural elements P1 to P9 found in nuclear group I introns (see Michel and Westhof 1990; Bhattacharya et al. 1994; Golden et al. 1998). A total of 219 aligned positions were selected for the final data set (alignment available as Supplementary Material online). A 50% majority rule consensus intron tree was inferred using MrBayes and the GTR + I +  $\Gamma$  model that was identified with Modeltest version 3.06 (Posada and Crandall 1998) as the best-fit model. MCMCMC from a random starting tree was initiated in the Bayesian inference and run for 5,000,000 generations with trees sampled every 100th generation. The first 10,000 trees were discarded as “burn-in,” and the subsequent 40,000 were used to calculate the consensus tree. The stability of nodes in the intron tree was also estimated with bootstrap analysis (1,000 replications) using neighbor-joining and distance matrices calculated with the GTR + I +  $\Gamma$  model and the LogDet transformation (using PAUP\*). The sequence of the zygomycete *Coemansia mojavensis* was used to root the tree of ascomycete and basidiomycete S943 introns.

#### Comparison of SSU rDNA and S943 Group I Intron Trees

To address intron inheritance, we inferred an SSU rDNA tree from the same taxa that contained the S943 group I intron to compare directly the intron and host trees. The rDNA alignment of 1,912 sites was prepared through juxtaposition of secondary structure elements to aid delimitating the ambiguous regions (Kjer 1995; Lutzoni

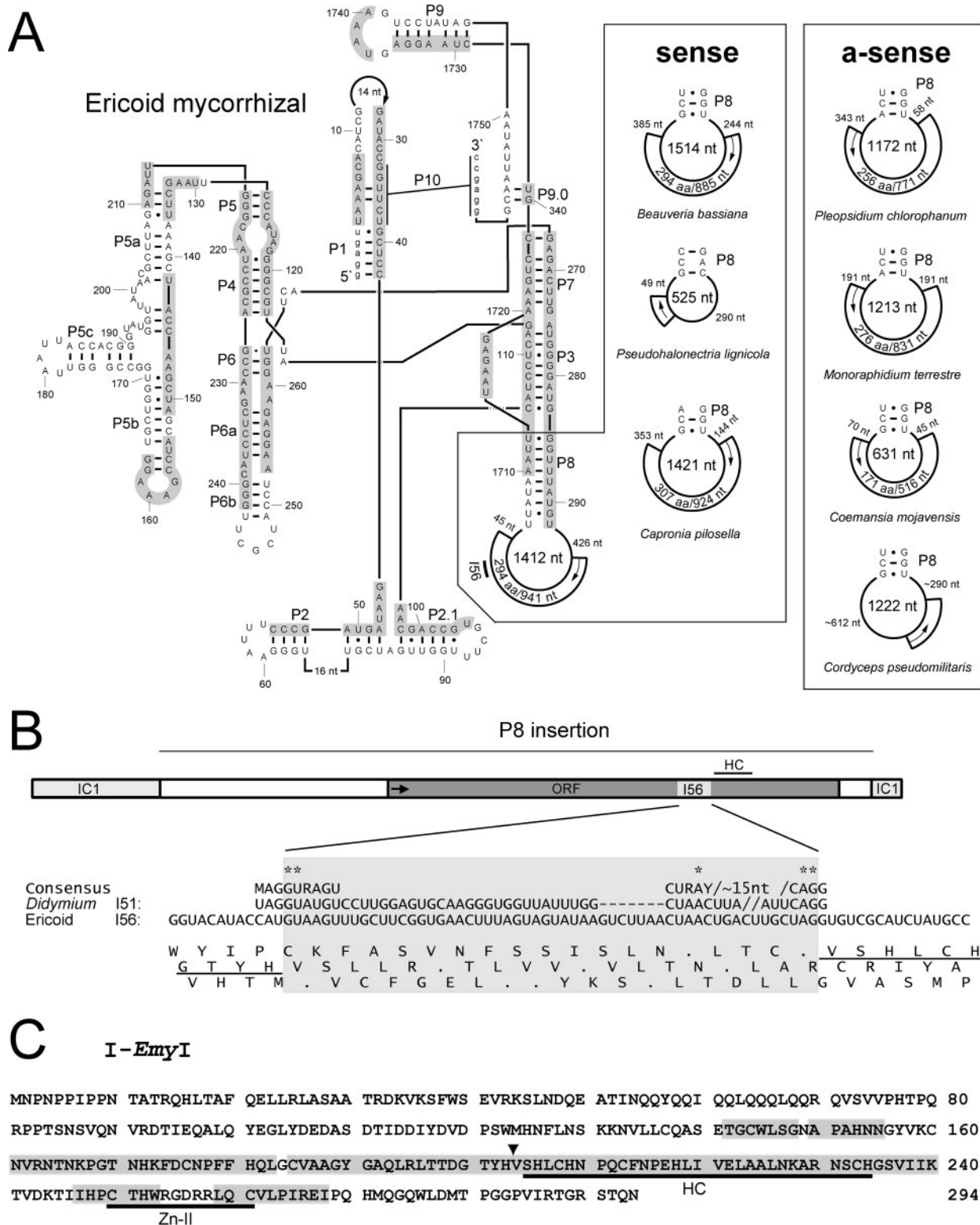


FIG. 1.—Structure and organization of the S943 group I intron from the fungal ericoid mycorrhizal PSIV isolate. (A) Secondary structure of the ericoid mycorrhizal S943 group I intron. Paired elements (P1 to P10) and numbering of every 10th nucleotide positions in the intron are marked on the structure. Shaded sequence positions were used to infer S943 intron phylogeny (see sequence alignment in Supplementary Material online). Homing endonuclease gene (HEG) insertions located in the P8 structural element of S943 group I introns of different species are boxed. Arrows indicate the orientation of the HEG. Four of the HEG insertions are encoded by the same strand as the group I ribozyme (sense), whereas four are found on the strand complementary to that of the group I ribozyme, denoted as the antisense (a-sense) strand. The I56 spliceosomal intron-like sequence is indicated in the ericoid mycorrhizal S943 group I intron. (B) Linear organization of the P8 insertion in the ericoid mycorrhizal S943 group I intron. The arrow shows the ORF start and orientation. The putative I56 spliceosomal intron is shaded and the region encoding the His-Cys box is marked (HC). I56 has strong identity to the spliceosomal intron consensus sequence (Lopez and Séraphin 1999) and to a spliceosomal intron (I51) found in the S956 group I

et al. 2000). The Bayesian analysis (GTR + I +  $\Gamma$  model) was restricted to the non-ambiguously aligned portions (1,387 sites) of this alignment and performed as described above for the S943 data set, to generate a 50% majority rule consensus tree. We then used a weighted maximum-parsimony analysis to compare the level of resolution and support gained when integrating phylogenetic signal from ambiguously aligned regions, with support values derived from the Bayesian MCMCMC analysis. In this analysis, a total of 43 ambiguous rDNA regions were delimited, of which eight were excluded from the analyses, 30 were treated with INAASE version 2.3b (Lutzoni et al. 2000 [http://www.morag.com/lutzoni/download.shtml#INAASE]), and five were treated with Arc version 1.5 (Miadlikowska et al. 2003 [program available by request to Frank Kauff or FL]). The unambiguously aligned portions of the data matrix were subjected to a specific symmetric step matrix that accounts for the frequency of changes between all possible character states (four nucleotides and gaps as a fifth character state). The frequency matrix was then converted to a matrix of cost of changes using the negative natural algorithm (Felsenstein 1981; Wheeler 1990). The bootstrap parsimony analysis was run using 1,000 pseudoreplicates with four random addition sequences, TBR branch swapping, and a reconnection limit of 8. Nexus files of the weighted and unweighted SSU rDNA data sets are available as Supplementary Material online. The sequence of the zygomycete *Coemansia braziliensis* (the rDNA sequence of *C. mojavensis* is not available) was used to root the tree of ascomycete and basidiomycete SSU rDNA genes.

To compare phylogenetic positions in the intron and host trees, we used the reciprocal 70% bootstrap support criterion proposed by Mason-Gamer and Kellogg (1996). Under this criterion, if the same set of terminal taxa receive bootstrap support of 70% or more for a monophyletic relationship in a data partition and is not monophyletic with support values of 70% or more in a different data partition, then we interpret this as a topological conflict and as evidence supporting a lateral transfer.

#### Phylogenetic Analysis of HEG-Containing Group I Introns

All nuclear introns that contain HEGs were included in an alignment (206 nt) and analyzed with the DNA maximum-likelihood method. The K80 (K2P) +  $\Gamma$  model, identified with Modeltest as the best-fit model, was used in this reconstruction (alignment available as Supplementary Material online). Starting trees were built stepwise (10 random additions) and optimized by TBR. The stability of nodes in the intron maximum-likelihood tree was estimated with bootstrap analysis (1,000 replications) using the minimum-evolution method and distance matrices calculated with the K80 (K2P) +  $\Gamma$  model and the LogDet

transformation. Ten heuristic searches with random-addition of sequences and TBR branch rearrangements were used to find each optimal bootstrap minimum evolution tree. We also used MrBayes (GTR + I +  $\Gamma$  model), as described above for the S943 data set, to calculate posterior probabilities of the nodes resolved in the intron maximum-likelihood tree. The group IE introns were used to root these trees.

## Results and Discussion

### Group I Introns at Position 943 in the SSU rDNA

The S943 introns, together with introns in positions S516, S1506, and S1516, account for more than 50% of all identified group I introns in nuclear rRNA genes with 155, 174, 213, and 166 introns, respectively (as of January 15, 2003 [Cannone et al. 2002]). In this current work, we have studied the S943 introns. The overall structure and organization of all S943 introns are similar with a group IC1 type ribozyme folding. Four S943 introns have previously been reported to encode intact or pseudogene HEGs. These HEGs encode the putative homing ENases I-*EmyI* from an unidentified ericoid mycorrhizal fungus (Perotto et al. 2000), I-*MteI* from *Monoraphidium terrestre* (Chlorophyceae [Krienitz et al. 2001; see Haugen, De Jonckheere, and Johansen 2002]), I-*CmoI* from *Coemansia mojavensis* (Zygomycetes [Tanabe, Yokota, and Sugiyama 2002]), and I-*BbaI* from *Beauveria bassiana* (Sordariomycetes [Yokoyama, Yamagishi, and Hara 2002]). The HEGs are all coded by sequence extensions located in the P8 element of their respective S943 introns (fig. 1A). We compiled a list of 81 S943 introns (table 2 in Supplementary Material online), of which 15 have insertions greater than 50 bp in P8, suggesting an HEG or HEG remnant. Analysis of these insertions (see below) identified two additional pseudogene HEGs. Finally, we sequenced two new S943 introns, which also encode HEGs in P8. Therefore, we arrive at a total set of 81 introns, of which eight have inferred functional or nonfunctional HEGs. The two fungal S943 introns with HEGs reported in this work are from *Pleopsidium chlorophanum* and *Capronia pilosella* (Pch.S943 and Cpi.S943, according to the latest intron nomenclature [Johansen and Haugen 2001]).

### HEGs in S943 Introns Are Often Inactivated

Of the four previously identified HEGs in S943 introns, three are inferred full length (I-*EmyI*, I-*MteI*, and I-*CmoI*) and one (I-*BbaI*) has a premature stop codon near the C-terminus (pseudogene). A C-terminal sequence very similar to that of I-*CmoI* can be reconstructed by ignoring the stop codon located after the last residue (294) of the I-*BbaI* HEG. The newly sequenced Pch.S943 and Cpi.S943 introns appear to contain full-length HEGs encoding

←

intron-encoded HEG of *Didymium iridis*. Consensus nucleotide positions corresponding to A or C (M), A or G (R), and C or U (Y) are shown. The asterisks indicate invariable sites in the consensus sequence. The paired slashes (//) in the I51 sequence indicate 4 nt that are not shown in this figure. (C) The amino acid sequence of the I-*EmyI* homing endonuclease encoded by the ericoid mycorrhizal S943 group I intron. The His-Cys box (HC) and the C-terminal zinc-II motif (Zn-II) are underlined. The arrowhead marks the position of I56. The shaded sequence positions were used to infer the HEG phylogeny (i.e., see sequence alignment in Supplementary Material online).

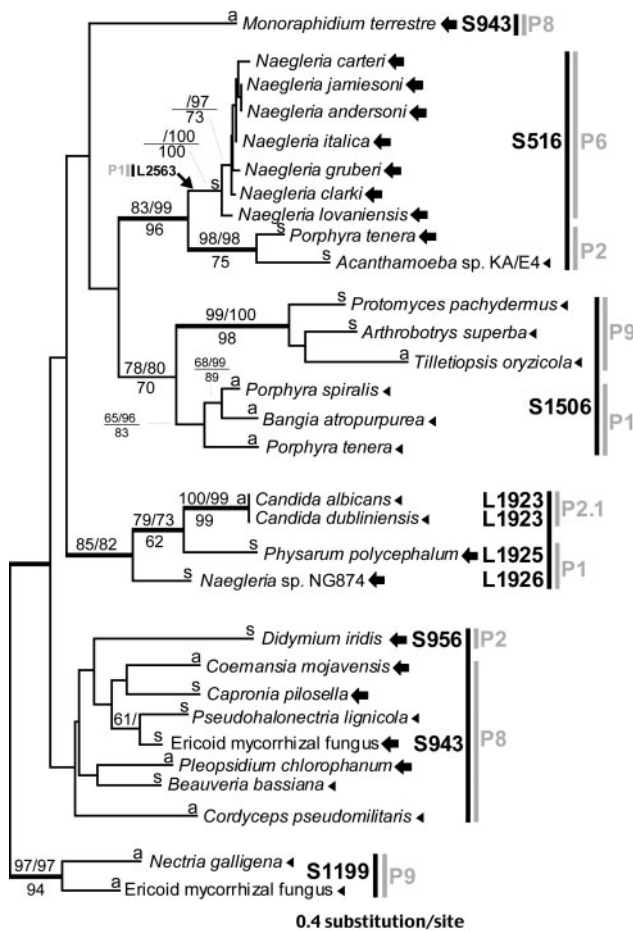


FIG. 2.—Phylogeny of nuclear group I intron-encoded homing endonucleases (ENases) based on 119 aligned amino acids. The tree was generated using the protein maximum-likelihood (proml) method (JTT +  $\Gamma$  model). The bootstrap values above the internal nodes on the left of the slash marks were inferred with the proml method, whereas the bootstrap values on the right of the slash marks were inferred with a minimum-evolution analysis (Poisson model). The bootstrap values under the branches were inferred with an unweighted maximum-parsimony analysis. Only bootstrap values of 60% or more are shown. The thick branches denote greater than 95% posterior probability for groups to the right resulting from a Bayesian inference (WAG +  $\Gamma$  model). The sense (s) or antisense (a) orientation of the HEGs compared with the ribosomal RNA gene are indicated on the branches. Inferred functional (large filled arrows) and nonfunctional (arrowheads) ENases are shown. Thick black vertical lines indicate clusters of related ENases (intron insertion site shown). The gray text (P1 to P9) and vertical lines mark the HEG position in the ribozyme secondary structure. The branch lengths are proportional to the number of substitutions per site (see scale in figure).

putative homing endonucleases of 256 aa and 307 aa, respectively. The proteins were denoted *I-PchI* and *I-CpiI*, according to the current nomenclature (Belfort and Roberts 1997). *I-PchI* is encoded on the antisense (a-sense) strand of rDNA (fig. 1A) and contains an intact His-Cys box and a Zn-II domain. *I-CpiI* also contains a conserved His-Cys box and Zn-II motif but is encoded on the sense strand (fig. 1A). Analysis of the P8 insertions in our compilation identified two more pseudogenes. The *Pseudohalonectria lignicola* intron was found to encode a Zn-II-like motif very similar to that found in other S943 homing ENases on the a-sense strand. However, we were unable to identify

a longer ORF or a complete His-Cys box. Similarly, the *Cordyceps pseudomilitaris* intron (Cps.S943) was found to no longer encode a recognizable ORF on any of the six frames translated from the P8 sequence. However, by introducing four frameshifts, a theoretical protein sequence of about 123 aa resembling a homing ENase with both the conserved His-Cys box and Zn-II domains was reconstructed in silico. Therefore, of the eight HEGs in S943 introns, five encode full-length HEGs and three encode pseudogenes. In addition, seven S943 introns have insertions in P8 that do not encode ORFs but might be HEG remnants. These findings are consistent with the idea that HEGs in group I introns are unstable and tend to be lost over time (e.g., Cho et al. 1998; Haugen et al. 1999; Goddard and Burt 1999; Foley, Bruttin, and Brüssow 2000; Müller et al. 2001; Bhattacharya, Friedl, and Helms 2002; Nozaki et al. 2002).

#### The *I-EmyI* Endonuclease Gene Is Interrupted by a Spliceosomal Intron-like Sequence

The ericoid mycorrhizal fungal isolate PSIV contains a large group I intron of 1,759 nt at position S943 (Emy.S943) with a HEG-like sequence inserted in P8 (fig. 1A [Perotto et al. 2000]). The P8 extension encodes a protein of length 30 aa with a His-Cys box motif on the sense strand. During the computer-repair approach to this protein-coding sequence beyond the stop codon located two codons upstream of the His-Cys box, we discovered a 56-nt region with strong similarity to conserved regions of yeast spliceosomal introns (Lopez and Séraphin 1999) and to the 51-nt spliceosomal intron from the *I-Diri* HEG (Vader, Nielsen, and Johansen. 1999; fig. 1B). By removing the putative intron sequence from the Emy.S943 HEG, we were able to identify a protein of length 294 aa (*I-EmyI* [fig. 1C]).

#### Phylogeny of His-Cys Box Homing Endonucleases

By using the conserved ribozyme core sequence, the phylogeny and origin of group I introns has been addressed in recent studies (e.g., Hibbett 1996; Bhattacharya 1998; Perotto et al. 2000; Friedl et al. 2000; Müller et al. 2001; Nikoh and Fukatsu 2001; Bhattacharya, Friedl, and Helms 2002). However, very little is known about the phylogeny of intron-encoded nuclear homing endonucleases and how the presence of these ORFs relates to the evolutionary history of the group I introns in which they reside. Using the amino acid sequence of previously reported nuclear homing endonucleases and those reported in this work (see *Materials and Methods*), a proml tree was inferred using the 119-aa alignment (see fig. 1C and *Materials and Methods* for details). The four different phylogenetic methods used in this analysis all provided comparable levels of support for monophyletic HEG lineages in the proml tree.

The homing endonuclease tree (fig. 2) shows that the S516 endonucleases from the distantly related *Naegleria*, *Porphyr*a, and *Acanthamoeba* (i.e., see Baldauf et al. 2000) form a monophyletic group with moderate to high bootstrap support. The S1506 ENases from three closely related red algae in the Bangiophycidae and three fungal

species (one basidiomycete and two ascomycetes) are also grouped together with moderate support. Furthermore, the two S1199 proteins from *Nectria galligena* and ericoid mycelia form a well-supported clade. The *Monoraphidium terrestre* S943 HEG does not group with the fungal HEGs at this site, but its virtually unsupported position in the tree does not allow us to infer the phylogenetic history of the sequence. The single L2563 HEG from *Naegleria morganensis* could also not be placed with bootstrap support in any clade, and its presence significantly weakened the support for other HEG lineages. Bootstrap support for the monophyly of the S516 and L1923 to L1926 HEG clades increased from 80% (including L2563) and 75% to 99% and 82%, respectively, in the minimum-evolution analysis when the L2563 sequence was excluded from the data set. Therefore, the L2563 HEG sequence was removed from the analyses, and its computer-predicted (although unlikely) phylogenetic position is indicated with an arrow in figure 2.

Although the data are relatively sparse, our data suggest that homing ENases from introns at homologous sites (e.g., S516 and S1506) are often closely related, although they may reside in evolutionarily distantly related host organisms (e.g., see also Pellenz et al. [2002]). More interesting is the finding that homing ENases encoded by introns from neighboring rDNA sites group together. The two *Candida* HEGs I-*CalI* and I-*CduI* (L1923 [fungi]), the I-*PpoI* (L1925 [Mycetozoa]), and the I-*NaeI* (L1926 [Heterolobosea]) ENases are encoded by introns that are located within 3 nt of each other in the LSU rDNA and form a clade with moderate to high bootstrap and significant Bayesian support (fig. 2). Similarly, the S956 ENase I-*DirI* from the Mycetozoa *Didymium iridis* is nested within the fungal S943 clade (albeit without bootstrap or Bayesian support).

#### Phylogeny of S943 Group I Introns

To understand how the distribution of homing ENases inserted into S943 group I introns of fungi correlates with the phylogeny of this intron, we generated, using Bayesian inference, a 50% majority rule consensus tree of the S943 intron RNA sequences (fig. 3). Introns containing inferred active ENases, ENase pseudogenes, or sequence insertions over 50 bp in length in P8 were mapped on the tree. This distribution clearly shows that HEGs, or putative HEG remnants, are widely distributed among S943 introns. The ascomycete introns (except in the Taphrinomycotina and Saccharomycotina) are particularly rich in HEG-like sequences with three inferred full-length HEGs (marked with arrows [Ericoid mycorrhizal HEG not shown; see table 2 in Supplementary Material online]) and three pseudogenes (arrowheads). In addition, six ascomycete introns contain sequence insertions over 50 nt in length in P8 (open arrows), suggesting that these introns once contained HEGs at the homologous position. The widespread and sporadic distribution of these sequences within the Pezizomycotina (Ascomycota) may be interpreted as evidence for the repeated horizontal movement of the HEG and/or intron-HEG combination with the rapid subsequent loss of the HEG. Alternatively,

the S943 HEG may be an ancestral sequence in these fungi (i.e., the outgroup zygomycete contains an active HEG), and it has been vertically inherited in three lineages and lost from the majority. Finally, the present distribution may represent some combination of vertical ancestry and lateral transfer across the fungal tree for the S943 intron and HEG either as a unit or as independent elements. Comparisons of intron and host trees are required to address this issue (see below). In any case, our S943 intron analyses show that HEG inactivation and loss appears to be pervasive and might result after all the rDNA alleles in a population were fixed for the intron (through homing) and the ENase would then lose its function (e.g., as suggested by Goddard and Burt [1999]).

#### Comparison of SSU rDNA and S943 Intron Trees

To address the mobility of fungal S943 introns, we compared SSU rDNA and S943 intron trees from equivalent taxa (fig. 3). The low resolution in the intron tree does not allow us to make any strong conclusions, but nevertheless the data give us some clues into intron mobility. In general, the juxtaposition of the phylogenies suggests vertical ancestry of the fungal S943 introns. A number of sequences are “misplaced” in the intron tree (e.g., *Graphium basitruncatum* and *G. penicillioides*) with the assumption that the SSU rDNA tree is the true phylogeny, however only one conflict meets the reciprocal 70% bootstrap support criterion (Mason-Gamer and Kellogg 1996). This is the position of *Beauveria bassiana* as sister to *Isaria japonica*, *Paecilomyces tenuipes*, and *P. cicadae* in the intron tree, whereas *B. bassiana* shows a well-supported sister group relationship to *Cordyceps brongniartii* in the rDNA tree (fig. 3). This is likely a case of intron + HEG mobility because the *B. bassiana* intron encodes a HEG sequence (i.e., inactivated by a single mutation). More compelling is the general congruence of the rDNA and intron phylogenies over much of the well-supported regions of the trees (except *B. bassiana*) with no evidence of recent lateral transfers. The lack of evidence of recent horizontal transfers does not, however, exclude the possibility that such events have taken place.

#### HEG Phylogeny Versus Intron Phylogeny

Next, we wanted to address how the phylogeny of all known ribozymes that encode the HEG sequences correlates with the phylogeny of the nuclear group I intron-encoded homing ENases. A 50% majority rule consensus of the 18 intron trees of equal likelihood that were identified with PAUP\* was juxtaposed to the ORF tree shown in figure 2 (see fig. 4). In this analysis, similar intron and ORF tree topologies were taken to suggest a common evolutionary history of the intron and HEG sequences, either through vertical inheritance of the intron/HEG combination or through horizontal transfer of the intron/HEG as a unit. In contrast, incongruence of the two trees was taken as evidence for independent evolutionary histories of the group I introns and the HEGs. Other issues that we considered were the position of the HEG in the ribozyme secondary structure and the strand on which the HEG was coded (fig. 2).

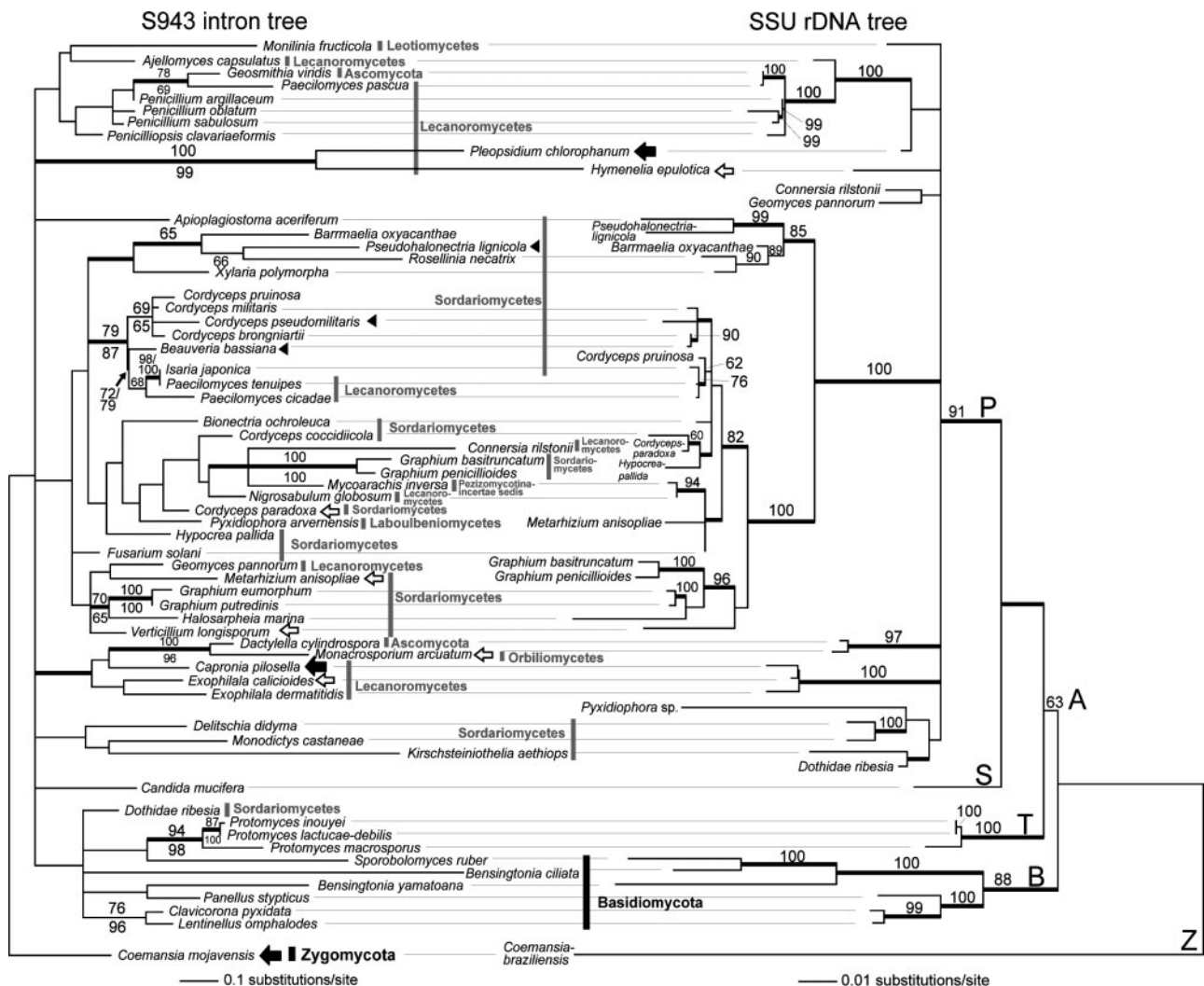


FIG. 3.—Comparison of S943 group I intron and SSU rDNA trees. The intron phylogeny (left in figure) is a 50% majority rule consensus tree generated using an alignment of 219 nt and Bayesian inference (GTR + I +  $\Gamma$  model). The thick branches denote greater than 95% posterior probability for groups to the right. The bootstrap values above the internal nodes were inferred from a neighbor-joining-GTR + I +  $\Gamma$  analysis, whereas the bootstrap values under the nodes were inferred from a second neighbor-joining analysis using LogDet distances. The arrows indicate sequence insertions in the P8 element corresponding to inferred full-length HEGs (large filled arrows), inactivated HEGs (arrowheads), or sequence insertions not encoding a putative HEG ORF but being over 50 bp in length (open arrows). The SSU rDNA tree (right in figure) is a 50% majority rule consensus tree generated using an alignment of 1387 nt and Bayesian inference (GTR + I +  $\Gamma$  model). The thick branches denote greater than 95% posterior probability for groups to the left. The bootstrap values at the internal nodes were inferred from a weighted maximum-parsimony analysis. Only bootstrap values of 60% or more are shown in these trees. Classification within the Ascomycota orders follows GenBank, and the classification above the ordinal level follows Taylor et al. (2003). The letters Z, B, A, T, S, and P denote Zygomycota, Basidiomycota, Ascomycota, Taphrinomycotina, Saccharomycotina, and Pezizomycotina, respectively. The branch lengths are proportional to the number of substitutions per site (see scales in figure).

Using these criteria, the most convincing cases of intron-HEG coinheritance are the two fungal S1199 introns and the S516 intron/HEGs of several *Naegleria* species. The S516 intron in closely related *Naegleria* species is most likely the result of vertical evolution of the intron and the HEG. With the S1199 introns, it is, however, unclear how closely related the two fungal hosts are (unidentified Ericoid mycorrhizal fungus and *Nectria galligena*), thus making it difficult to discriminate between vertical inheritance and horizontal transfer. For the S1506 introns, the data suggest two sets of vertically inherited introns, one in ascomycetes and one in algae. The positioning of the S1506 HEGs in different RNA structural elements and/or on opposite strands, suggests that there

were at least two horizontal transfers of the intron HEG, one between an ascomycete and a basidiomycete (*P. pachydermus* or *A. superba* [sense strand] and *T. oryzicola* [a-sense strand], respectively) and one between fungi (ascomycete or basidiomycete) and red algae (fungal P9 domain, red algal P1). The first example may, however, be vertical HEG inheritance if strand switching is not necessarily associated with HEG movement.

The single example of a clear incongruence between group I intron and HEG trees is boxed in figure 4. In this case, several phylogenetic methods show that the ENases of two *Candida* species, *Physarum polycephalum* and *Naegleria* sp. NG874, are encoded by neighboring introns (L1923 to L1926) and branch together on the ORF tree.



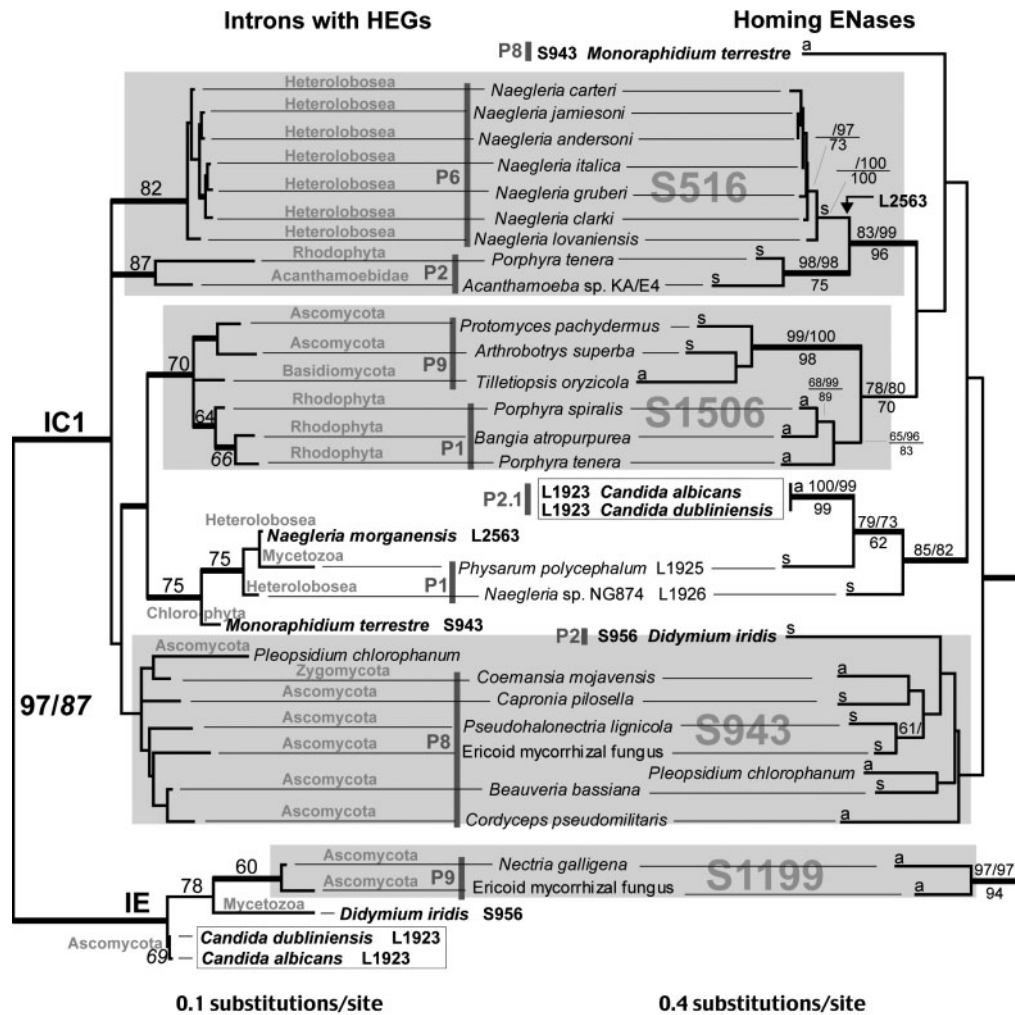


FIG. 4.—Comparison of phylogenies of HEG-containing nuclear group I introns and the corresponding homing ENases. Shown on the right is a mirror image of the ENase tree from figure 2. The intron tree (left in figure) was generated using the maximum-likelihood method (K80 [K2P] +  $\Gamma$  model). The bootstrap values above the internal nodes of the latter tree were inferred from a minimum-evolution analysis (K80 [K2P] +  $\Gamma$  model), whereas the bootstrap values under the nodes were inferred from a second minimum-evolution analysis using LogDet distances. Only bootstrap values of 60% or more are shown. The thick branches denote greater than 95% posterior probability for groups to the right resulting from a Bayesian inference (GTR + I +  $\Gamma$  model). The deep separation of introns belonging to the IC1 or IE subgroups is shown on the tree (large IC1 and IE text). Shaded areas highlight groups of introns from the same intron insertion site (shown in large gray text) and regions where the two tree topologies show strong similarity. A clear case of tree incongruence is boxed and shown in bold text. The thick, gray vertical lines and text (P1 to P9) indicate HEG position in the ribozyme. The branch lengths are proportional to the number of substitutions per site (see scales in figure).

However, whereas the *P. polycephalum* and *Naegleria* sp. NG874 ribozymes belong to the group IC1 subclass, the *Candida* ribozymes belong to the distantly related group IE subclass (Suh, Jones, and Blackwell 1999). Clustering of the L1923 to L1926 ORFs has strong bootstrap and significant Bayesian support, as well as the deep separation of the IC1 and IE ribozyme classes. The S943 HEG is also a potential case of HEG mobility because of the extensive HEG strand switching (“a” versus “s” in figs. 2 and 4) that has occurred in this clade.

How do HEGs move from one intron to another? Present understanding suggests that HEG insertion into introns occurs through illegitimate double-strand, break-mediated gene conversion (e.g., Loizos, Tillier, and Belfort 1994; Parker, Belisle, and Belfort 1999). HEG movement is generally limited to insertion into homologous introns. This reflects the fact that when HEGs invade introns, the

organism is only protected from the ENase if the intron interrupts the ENase recognition sequence (Loizos, Tillier, and Belfort 1994). Such a scenario favors ENase invasion into introns at homologous or, as we also observe, neighboring (i.e., L1923 to L1926 and S943 to S956) heterologous rDNA sites that also disrupt the sequence recognized by the ENase. Once the ENase has been established in the homologous or heterologous intron, it can mobilize that intron into its natural site through homing.

#### Conclusion

In this work, we present seven previously unknown or unidentified nuclear HEGs resulting in a total of 22 different nuclear sequences for our phylogenetic analyses. We studied the evolutionary history of group I introns

inserted into the S943 site in fungi and algae to address the following specific questions: (1) Does the S943 HEG distribution primarily favor vertical ancestry of the gene, or are these sequences frequently laterally transferred between introns in the same or different species? (2) Are the S943 introns frequently laterally transferred among fungi? (3) How frequently are HEGs transferred to ectopic sites, and if so, in which sites are they fixed in nature? Our analyses provide answers to these questions by showing first that HEGs are widely distributed on the S943 intron tree. Many of the S943 HEGs are nonfunctional or remnant ORFs. This finding plus the observation that S943 HEGs are inserted in different orientations and have highly uncertain phylogenetic relationships, suggests that they may move horizontally into introns and are then lost. In answer to the second question, detailed phylogenetic analyses of the S943 introns and SSU rDNAs failed (except for the *B. bassiana* case) to provide support for pervasive intron lateral transfer. To address the third and last question we aligned all available sequences of His-Cys box HEG proteins and reconstructed their phylogeny. This tree shows that HEGs encoded by introns in the same position are related and that HEGs in the same intron insertion site can be inserted into different intron structural elements or orientations, suggesting that HEGs can reinvade the same intron or other introns inserted in the same position. The HEG protein tree also shows one case in which ENases from three neighboring (ectopic) sites are closely related. Interestingly, the introns in which they reside belong to two distantly related subgroups (IC1 and IE). This finding again suggests that HEG mobility is independent of the intron and can result in HEG insertion into distantly related introns in neighboring sites. Such a scenario is favored because an intron that is inserted only a few nucleotides from the original insertion site will disrupt the sequence recognized by the ENase and thus protect the DNA against endonuclease activity. A HEG inserted into a neighboring site will also support homing of the new intron host (in the neighboring site) because the DSBR pathway of homing includes coconversion of flanking exon sequences (Dujon 1989; Belfort and Perlman 1995).

### Supplementary Material

The HEG amino acid, S943 group I intron, SSU rDNA alignments, and table 2 are available on the MBE Web site.

### Acknowledgments

This work was supported by grants DEB 01-0774 and MCB 01-10252 awarded to D.B. from the National Science Foundation, a grant from The Norwegian Research Council to P.H., and an A. W. Mellon training grant from Duke University and a Grants-in-Aid of Research from Sigma-Xi to V.R. We would like to thank Sabine Huhndorf for kindly providing a culture of *Capronia pilosella* and two anonymous reviewers for their helpful comments.

### Literature Cited

- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**:972–977.
- Belfort, M., and P. S. Perlman. 1995. Mechanisms of intron mobility. *J. Biol. Chem.* **270**:30237–30240.
- Belfort, M., and R. J. Roberts. 1997. Homing endonucleases: keeping the house in order. *Nucleic Acids Res.* **25**:3379–3388.
- Benson, D.A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2003. GenBank. *Nucleic Acids Res.* **31**:23–27.
- Bhattacharya, D. 1998. The origin and evolution of protist group I introns. *Protist* **149**:113–122.
- Bhattacharya, D., J. J. Cannone, and R. R. Gutell. 2001. Group I intron lateral transfer between red and brown algal ribosomal RNA. *Curr. Genet.* **40**:82–90.
- Bhattacharya, D., T. Friedl, and G. Helms. 2002. Vertical evolution and intragenic spread of lichen-fungal group I introns. *J. Mol. Evol.* **55**:74–84.
- Bhattacharya, D., B. Surek, M. Ruesing, and M. Melkonian. 1994. Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). *Proc. Natl. Acad. Sci. USA* **91**:9916–9920.
- Cannone, J. J., S. Subramanian, M. N. Schnare et al. (14 co-authors). 2002. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**:1–31.
- Cech, T. R. 1990. Self-splicing of group I introns. *Annu. Rev. Biochem.* **59**:543–568.
- Chevalier, B. S., and B. L. Stoddard. 2001. Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. *Nucleic Acids Res.* **29**:3757–3774.
- Cho, Y., Y. L. Qiu, P. Kuhlman, and J. D. Palmer. 1998. Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. USA* **95**:14244–14249.
- Dujon, B. 1989. Group I introns as mobile genetic elements: facts and mechanistic speculations—a review. *Gene* **82**:91–114.
- Felsenstein, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16**:183–196.
- . 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- . 2002. PHYLIP (phylogeny inference package). Version 3.6a3. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.
- Flick, K. E., M. S. Jurica, R. J. Monnat Jr., and B. L. Stoddard. 1998. DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-*PpoI*. *Nature* **394**:96–101.
- Foley, S., A. Bruttin, and H. Brüssow. 2000. Widespread distribution of a group I intron and its three deletion derivatives in the lysin gene of *Streptococcus thermophilus* bacteriophages. *J. Virol.* **74**:611–618.
- Friedl, T., A. Besendahl, P. Pfeiffer, and D. Bhattacharya. 2000. The distribution of group I introns in lichen algae suggests that lichenization facilitates intron lateral transfer. *Mol. Phylogenet. Evol.* **14**:342–352.
- Gargas, A., and J. W. Taylor. 1992. Polymerase chain reaction (PCR) primers for amplifying and sequencing 18S rDNA from lichenized fungi. *Mycologia* **84**:589–592.
- Goddard, M. R., and A. Burt. 1999. Recurrent invasion and extinction of a selfish gene. *Proc. Natl. Acad. Sci. USA* **96**:13880–13885.
- Golden, B. L., A. R. Gooding, E. R. Podell, and T. R. Cech. 1998. A preorganized active site in the crystal structure of the *Tetrahymena* ribozyme. *Science* **282**:259–264.

- Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**:95–98.
- Haugen, P., J. F. De Jonckheere, and S. Johansen. 2002. Characterization of the self-splicing products of two complex *Naegleria* LSU rDNA group I introns containing homing endonuclease genes. *Eur. J. Biochem.* **269**:1641–1649.
- Haugen, P., V. A. Huss, H. Nielsen, and S. Johansen. 1999. Complex group-I introns in nuclear SSU rDNA of red and green algae: evidence of homing-endonuclease pseudogenes in the Bangiophyceae. *Curr. Genet.* **36**:345–353.
- Hibbett, D. S. 1996. Phylogenetic evidence for horizontal transmission of group I introns in the nuclear ribosomal DNA of mushroom-forming fungi. *Mol. Biol. Evol.* **13**:903–917.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17**:754–755.
- Johansen, S., and P. Haugen. 2001. A new nomenclature of group I introns in ribosomal DNA. *RNA* **7**:935–936.
- Jones, D., W. Taylor, and J. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**:275–282.
- Kauff, F., and F. Lutzoni. 2002. Phylogeny of the Gyalectales and Ostropales (Ascomycota, Fungi): among and within order relationships based on nuclear ribosomal RNA small and large subunits. *Mol. Phylogenet. Evol.* **25**:138.
- Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* **4**:314–330.
- Krienitz, L., J. Ustinova, T. Friedl, and V. A. R. Huss. 2001. Traditional generic concepts versus 18S rRNA gene phylogeny in the green algal family Selenastraceae (Chlorophyceae, Chlorophyta). *J. Phycol.* **37**:852–865.
- Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **12**:1244–1245.
- Lambowitz, A. M., and M. Belfort. 1993. Introns as mobile genetic elements. *Annu. Rev. Biochem.* **62**:587–622.
- Loizos N., E. R. Tillier, and M. Belfort. 1994. Evolution of mobile group I introns: recognition of intron sequences by an intron-encoded endonuclease. *Proc. Natl. Acad. Sci. USA* **91**:11983–11987.
- 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., 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**:628–651.
- Mason-Gamer, R. J., and E. A. Kellogg. 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). *Syst. Biol.* **45**:524–545.
- Miadlikowska, J., F. Lutzoni, T. Goward, S. Zoller, and D. Posada. 2003. New approach to an old problem: incorporating signal from gap-rich regions of ITS and nrDNA large subunit into phylogenetic analyses to resolve the *Peltigera canina* species complex. *Mycologia* (in press).
- Michel, F., and E. Westhof. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**:585–610.
- Mota, E. M., and R. A. Collins. 1988. Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron. *Nature* **332**:654–656.
- Müller, K. M., J. J. Cannone, R. R. Gutell, and R. G. Sheath. 2001. A structural and phylogenetic analysis of the group IC1 introns in the order Bangiales (Rhodophyta). *Mol. Biol. Evol.* **18**:1654–1667.
- Nikoh, N., and T. Fukatsu. 2001. Evolutionary dynamics of multiple group I introns in nuclear ribosomal RNA genes of endoparasitic fungi of the genus *Cordyceps*. *Mol. Biol. Evol.* **18**:1631–1642.
- Nozaki, H., M. Takahara, A. Nakazawa, Y. Kita, T. Yamada, H. Takano, S. Kawano, and M. Kato. 2002. Evolution of *rbcL* group IA introns and intron open reading frames within the colonial Volvocales (Chlorophyceae). *Mol. Phylogenet. Evol.* **23**:326–338.
- Ogawa, S., K. Naito, K. Angata, T. Morio, H. Urushihara, and Y. Tanaka. 1997. A site-specific DNA endonuclease specified by one of two ORFs encoded by a group I intron in *Dictyostelium discoideum* mitochondrial DNA. *Gene* **191**:115–121.
- Parker, M. M., M. Belisle, and M. Belfort. 1999. Intron homing with limited exon homology: illegitimate double-strand-break repair in intron acquisition by phage T4. *Genetics* **153**:1513–1523.
- Pellenz, S., A. Harington, B. Dujon, K. Wolf, and B. Schafer. 2002. Characterization of the I-*SpomI* endonuclease from fission yeast: insights into the evolution of a group I intron-encoded homing endonuclease. *J. Mol. Evol.* **55**:302–313.
- Perotto, S., P. Nepote-Fus, L. Saletta, C. Bandi, and J. P. Young. 2000. A diverse population of introns in the nuclear ribosomal genes of ericoid mycorrhizal fungi includes elements with sequence similarity to endonuclease-coding genes. *Mol. Biol. Evol.* **17**:44–59.
- Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Roman, J., M. N. Rubin, and S. A. Woodson. 1999. Sequence specificity of in vivo reverse splicing of the *Tetrahymena* group I intron. *RNA* **5**:1–13.
- 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.
- Suh, S. O., K. G. Jones, and M. Blackwell. 1999. A group I intron in the nuclear small subunit rRNA gene of *Cryptodoxyla hypophloia*, an ascomycetous fungus: evidence for a new major class of group I introns. *J. Mol. Evol.* **48**:493–500.
- Swofford, D. L. 2002. PAUP\*: Phylogenetic analysis using parsimony (\*and other methods). Version 4.0b8. Sinauer Associates, Sunderland, Mass.
- Tanabe, Y., A. Yokota, and J. Sugiyama. 2002. Group I Introns from Zygomycota: evolutionary implications for the fungal IC1 intron subgroup. *J. Mol. Evol.* **54**:692–702.
- Taylor, J. W., J. Spatafora, K. O'Donnell, F. Lutzoni, T. James, D. S. Hibbett, D. Geiser, T. D. Bruns, and M. Blackwell. 2003. The kingdom Fungi. In J. Cracraft, M. J. Donoghue, eds. (in preparation).
- Vader, A., H. Nielsen, and S. Johansen. 1999. In vivo expression of the nucleolar group I intron-encoded I-*DirI* homing endonuclease involves the removal of a spliceosomal intron. *EMBO J.* **18**:1003–1013.
- Wheeler, W. C. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. *Cladistics* **6**:269–275.
- Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**:691–699.

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.

Yokoyama, E., K. Yamagishi, and A. Hara. 2002. Group I intron containing a putative homing endonuclease gene in small

subunit ribosomal DNA of *Beauveria bassiana* IFO 31676. *Mol. Biol. Evol.* **19**:2022–2025.

Thomas Eickbush, Associate Editor

Accepted August 19, 2003