Mitochondrial Intrinsic Open Reading Frames in Podospora:
Mobility and Consecutive Exonic Sequence Variations

Carole H. Sellem, Yves d’Aubenton-Carafa, Michèle Rossignol and Léon Belfort

Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette, France

Manuscript received July 17, 1995
Accepted for publication February 22, 1996

ABSTRACT

The mitochondrial genome of 25 wild-type strains belonging to three different species of the filamentous fungus Podospora was examined. Among the 15 optional sequences identified are two intrinsic reading frames, nad1-i4-orf1 and coxl-i7-orf2. We show that the presence of these sequences was strictly correlated with tightly clustered nucleotide substitutions in the adjacent exon. This correlation applies to the presence or absence of closely related open reading frames (ORFs), found at the same genetic locations, in all the Pyrenomycete genera examined. The recent gain of these optional ORFs in the evolution of the genus Podospora probably account for such sequence differences. In the homoplasmic progeny from heteroplasmic strains differing by the presence of these optional ORFs, nad1-i4-orf1 and coxl-i7-orf2 appeared highly invasive. Sequence comparisons in the nad1-i4 intron of various strains of the Pyrenomycete family led us to propose a scenario of its evolution that includes several events of loss and gain of intronic ORFs. These results strongly reinforce the idea that group I intronic ORFs are mobile elements and that their transfer, and concomitant modification of the adjacent exon, could participate in the modular evolution of mitochondrial genomes.

GROUP I introns are nonessential genetic elements. Most confer no apparent selective advantage at the individual level (SERAPHIN et al. 1987) but have spread in the course of evolution because endowed with invasive properties imparted by the product of the open reading frames (ORFs) they contain (LAMBOWITZ and BELFORT 1993). Their homing from intron-containing gene to cognate intronless allele is dependent upon this product, a site-specific endonuclease that creates a double-strand break at the exon-exon junction of the intronless allele (JACQUIER and DUJON 1985; MACREA-DIE et al. 1985; COLLEAUX et al. 1986).

The two distinct domains of mobile group I introns (catalytic core and endonuclease-encoding ORF) are supposed to have evolved independently (LAMBOWITZ and BELFORT 1993). The presence of related intronic ORFs in divergent organisms was already proposed to be the result of genetic exchanges involving the ORF sequence independently of the intron (MICHEL and DUJON 1986). Likewise, the presence in highly conserved introns of ORFs potentially encoding endonucleases that vary in both sequence and position (i.e., CUMMINGS and DOMENICO 1988; MOTA and COLLINS 1988; SHUB et al. 1988) also strongly suggests that the ORFs rather than the introns were the original mobile elements (BEL-PEDERSEN et al. 1990). Moreover, ORF-encoding, or potentially encoding, endonucleases similar to those found in group I introns have been identified in various other contexts such as in an archeal intron (DALGAARD et al. 1993) and as free standing sequences, inside (i.e., the inteins; COOPER and STEVENS 1995) or outside (SHARMA et al. 1992; PEL and GRIVEL 1993; PAQUIN et al. 1994) gene context. In some of these cases, the ORFs were described as optional and proposed to be recent acquisitions, but their mobility as independent genetic entities has only been demonstrated in the case of inteins (GIMBLE and THORNER 1992). Colonization of pre-existing group I introns by independent mobile ORFs is supported by a few examples of optional ORFs (MOTA and COLLINS 1988; BECHHOFER et al. 1994; BINISZKIEWICZ et al. 1994; DE JONCKHEERE and BROWN 1994).

Different mechanisms of acquisition have been proposed (LOIZOS et al. 1994; VADER et al. 1994), but no experimental evidence of mobility of such ORFs, from ORF-containing intron to cognate ORF-less allele, has been described.

The mitochondrial chromosome of the filamentous fungus Podospora anserina (P. an) contains numerous introns (up to 36 in the reference wild-type strain P. an(a)), some of which are optional (L. BELCOUR, unpublished results). Among the optional sequences are two ORFs potentially encoding highly specific DNA endonucleases, located within two group I introns: the fourth intron of the nad1 gene (nad1-i4) and the seventh intron of the coxl gene (coxl-i7), in the reference wild-type strain (CUMMINGS et al. 1990). Both introns have the peculiarity of containing either one (mono-orfic introns) or two (biorfic introns) ORFs depending on the wild-type isolate. We have examined the nad1 and coxl loci in other wild-type strains of the Pyrenomycete.
family: *P. comata* (*P. co*), *P. curvicolor* (*P. cu*) and *Sordaria macrospora* (*S. ma*). Three organizations were evidenced for both genes: intronless, mono- and biorfic. Sequence analysis revealed a systematic correlation between the presence/absence of the optional ORFs and the consensus sequence variation in the few adjacent nucleotides of the neighbouring exons (*nadl-i4* and *cox1-e8*). This correlation applies to the various organizations found for the equivalent loci in Neurospora (*Burger et al. 1982; Mota and Collins 1988; Field et al. 1989; Hawse et al. 1990*). We also took advantage of the favorable situation in *Podospora* to study the behavior of two optional intronic ORFs in heteroplasmons and demonstrated their efficient invasiveness. The data we present here led us to propose that these ORFs are recent acquisitions by the *Podospora* *nadl-i4* and *cox1-i7* introns and that their arrival was responsible for the modifications in the exonic sequences and thus contributed to mitochondrial genome evolution.

**MATERIALS AND METHODS**

**Strains:** *P. anserina* and *P. comata* are pseudohomothallic four-spored species that give some viable progenies when crossed one to each other. Most of the *P. anserina* wild-type strains used in this work (*A, B, C, D, E, F, H, M, N, R, S, U, V, W, X, Y, Z*) are French isolates provided by D. Marcoz. Two wild-type strains of *P. comata* were obtained from the ATCC: P. *co2* (no. 26491, Uganda) and P. *co3* (no. 36713, Venezuela), a colorless and totally sterile strain incompatible with any others. *P. coT* was a French isolate. *P. curvicolor* is a homothallic 128-spored species. *P. cuM* was given by J. Begeret. *P. cuL* and *P. cuV* were French isolates. *S. macrospora* strain was obtained from Dr. D. Zickler (ATCC no. 60255).

**Construction of isogenic strains:** *P. anserina* wild-type strains A, M and D (*P. anA, P. anM, P. anD*) were used as the male in crosses with the wild-type strain *P. anS*. In the progeny, the strains that did not present a 'barrage' with the *P. anS* strains were crossed again with *P. anS* as the male. Two cycles of backcrossing gave rise to strains containing a *P. anS*-like nucleus and either *P. anA, P. anM* or *P. anD* mitochondrial DNA. In the case of interspecific crosses, the *P. coT* wild-type strain was used as the male with *P. anS* as the female. Ten cycles of isogenization gave rise to strains containing *P. anS* mitochondria and rather pure *P. coT* nuclei. In all the cases, the parental strains used in heterokaryon construction were named according to the origin of their mitochondria.

**Construction of heterokaryons:** All the heterokaryons were constructed between pairs of strains of opposite mating type. Implants of growing mycelia of the ORF-less strains (recipients: *P. anM, P. anD* or *P. coT*) were placed on Petri dishes containing rich medium ~1 mm in front of the growing thallus of the ORF-containing strains (donors: *P. anA* or *P. anS*). When the mycelia had invaded the Petri dishes, they were exposed to light and fruiting bodies were allowed to develop. To study independent recombination events, a single ascus was collected from each heteroplastic cultures. Indeed, in *Podospora* heteroplastic mycelia were shown to produce only homoplasmic ascospores (*Belcour and Begel 1977*). Mycelia issued from 18 to 48 asci were further analyzed in each experiment.

**Analysis of mitochondrial DNA:** Mitochondrial DNAs were extracted from various *Podospora* wild-type, isogenic- or homoplasmic-derived strains by a minipreparation method (Lecellier and Silar 1994). Southern blot experiments were performed after *Hind*II or *Hinf*I digestions, on Hybond nylon membranes, using 32P-labeled synthetic oligonucleotides. *Ascothorax immersus* total DNA was a gift from J. L. Rossignol.

**Sequencing:** PCR amplifications were performed using pairs of synthetic oligonucleotides as primers. The fragments obtained were cloned in a *pMOSBlue* T-vector (Amersham) and the junctions sequenced using the PCR primers. The whole *P. co3* *nadl-i4* intron was sequenced on three independent clones to resolve the mistakes introduced by the Taq polymerase (Appligene); the *P. coT, P. anM* and *S. ma6nadl-i4* introns were cloned from restriction endonuclease fragments and sequenced either using specific oligonucleotide primers or using the nested deletion kit (Pharmacia) and the universal sequencing primers. The sequences have been entered into the EMBL bank under the number Z50030 for *P. coT* and Z50081 for *P. co3 nadl-i4* introns.

**Sequence data:** In all the strains of *P. anserina* and in strains of other species and genera, *nadl-i4* and *cox1-i7* equivalent introns, inserted at the same genomic location, are not necessarily the fourth or the seventh intron but will nevertheless be referred as *nadl-i4* and *cox1-i7* for simplicity. Neurospora and Aspergillus *nadl* and *cox1 mitochondrial* sequences, containing or not the *nadl-i4* and *cox1-i7* equivalent introns, were obtained from published data (see legend of Figure 1). Pairs of sequences were aligned using the BESTFIT program from the UWCGP package with a gap penalty of 5. Phylogenetic tree was built using DNADIST, FITCH and DRAWTREE programs from the PHYLIP package. The Lipman-Pearson algorithm (FASTA) was used for comparisons with EMBL (release 58) data banks.

**RESULTS**

The *nadl-i4* and *cox1-i7* introns in *P. anserina*: Mitochondrial DNA extracted from 17 wild-type strains of *P. anserina* was investigated by Southern blotting. The *nadl-i4* and *cox1-i7* introns, reported to be biorfic in the *P. anserina* wild-type strains A and S (*Cummings et al. 1988, 1989*) may present a different organization: in place of two ORFs (Figure 1, A and A') only one was found in several strains (Figure 1, B and B'). The *nadl-i4-orf1* and the *cox1-i7-orf2* are optional in *P. anserina*.

We have determined the nucleotide sequence of short DNA fragments (~200 bp) containing the *nadl-i4-i4* and the *cox1-i7-i8* junctions, after PCR amplifications of the mitochondrial DNA extracted from 10 wild-type strains of *P. anserina* (Figure 2, lines 4, 5 and 10, 11). For each region, all the strains displaying the biorfic intron presented a junction sequence identical to those already published (*Cummings et al. 1990*). The homologous regions in strains with a mono-orfic organization were found identical to each other, but in the exon adjacent to the optional ORF (the upstream one of *nadl-i4* and the downstream one of *cox1-i7*), a high level of divergence was observed with the corresponding sequence of the biorfic intron-containing gene. Indeed, nine and six substitutions are clustered in the last 28 nucleotides of the *nadl-i4* upstream exon and in the first 18 nucleotides of the *cox1-i7* downstream exon, respectively. This result was further extended to the 17 wild-type strains of *P. anserina* available in our labora-
Figure 1.—Organization of the nad1 and cox1 loci in various Pyrenomycete strains. The introns present at the same genetic location as P. anA nad1-i14 and cox1-i7 are either biorfic (A and A'), mono- or pseudomono-orfic (B and B' or D) introns. The structure in which the intron is absent from this position is also represented (C and C'). The 5'- and 3'-constitutive (△) and 3'-alternative (△) splice sites of the introns are indicated; their catalytic core is represented by a black dot (●). Exons are large boxes divided into several cross-hatched blocks representing the regions of polymorphic variations. The number of nucleotides in each block is indicated. Open bars indicate the upstream (ORF1) and the downstream (ORF2) intronic reading frames. The first ORFs are in frame with the upstream exon, show the two LAGLI-DADG dodecapeptides motifs, and loop out of the P1 and P2 secondary structures of the intron (nad1-i1-orf1 and cox1-i7-orf1, respectively). The second ORFs show either the GIY-YIG or LAGLI-DADG motifs (nad1-i1-orf2 and cox1-i7-orf2 respectively) laying in or downstream from the P9-1 secondary structure of the intron and require alternative splicing to be expressed. Bold arrows surround directly repeated exonic or intronic sequences bordering optional ORFs. The 17-bp (blocks 11+6) sequence, found downstream from the optional nad1-i1-orf1 (A), is an exact replica of the 3' end of the upstream exon bording the nad1-i14 mono-orfic intron (B). It contains a 6-bp invariant sequence corresponding to the 5' side of the P1 stem of the secondary structure of the intron. The 8-bp sequence found in the 5' end of the optional cox1-i7-orf2 is an exact replica of the downstream intron-exon junction; this replicated sequence contains the 3' side of the P9-0 and P10 stems in the secondary structure of the intron and thus provides an alternative 3'-splice site. The inset table indicates the organization of the introns for 17 P. anserina (A...Z), three P. comata (T, 2, 3) and three P. curvicaulis (M, L, V) strains. The intron organization for strains in other genera was either deduced by sequencing from S. macrospora strain 60255 (S. mnb; C. H. SELLEM, unpublished results) and A. immersus (A. im; C. H. SELLEM, unpublished results) or obtained from the literature: N. crassa strains 74A (N. cr7; BURGER et al. 1982; BURGER and WERNER 1985), Abbott 12 (N. crAb; HAWSE et al. 1990), Adiopodoume (N. crAd; FIELD et al. 1989), N. intermedia strain Varkan (N. inV; MOTA and COLLINS 1988), A. nidulans (A. ni; BROWN et al. 1983; WARING et al. 1984), Marchantia polymorpha (M. po; OTHA et al. 1993).

The wild-type strains P. anM and P. anD, respectively. It appears that the sequences absent from nad1-i14 and cox1-i7 mono-orfic introns contain the optional ORFs (391 and 325 encoded amino acids) and, respectively, 292 and 49 nucleotides downstream from their stop codon. In the case of nad1, over the 250 nucleotides of nad1-i4, the first 70 nucleotides of nad1-i5 and 1129 nucleotides of the mono-orfic intron, we found a perfect identity of sequence with the biorfic counterpart, excepted the nine positions we have mentioned and
Figure 2.—Alignments of exonic sequences. The 30 last nucleotides of the nad1-i4 upstream and of the 21 first nucleotides of the cox1-i7 downstream exons (uppercase letters) flanking introns (lowercase letters), containing the optional ORFs (O1* or O2*) or not (O1 or O2) or the corresponding sequences when the introns are absent (---), were compared. Data from P. anserina biotic (P. ass; Cummings et al. 1990) and mono-tropic nad1-i4 or cox1-i7 are given in reference (lines 4 and 5 or 10 and 11, respectively). The corresponding amino acid sequences are given and the codons rarely used in the mitochondrial genome of Podospora are labeled (*). Inset: Unrooted phylogenetic tree resulting from pairwise comparison of the nad1-i4 equivalent 30 last nucleotides among various species over the living kingdom. Distances were computed according to the Kimura two states parameters model assuming a transition/transversion ratio of 2. The tree was drawn using the Fitch-Margoliash criterion and the tree additive model with 10 runs of randomized input order of species. From the plant phylum, the chlorophyta Chlamydomonas reinhardtii (C. re), the liverwort Marchantia polymorpha (M. po), the water melon Citrullus lanatus (C. la) and the wheat Tritium aestivum (T. ae); from the animal phylum, the two nematoda Caenorhabditis elegans (C. el) and Ascaris suum (A. su), the blue mussel Mytilus edulis (M. ed), the lepidoptera Spodoptera frugiperda (S. fr) and the mammal Homo sapiens (H. sa); from the classical fungal phylum, devoid of nad1-i4 or nad1-i4 orf1; the Hymenomycete Trichophyton papulosum (T. po), the Pleurocytoma A. nidulans (A. ni), the Discomycete Aseobolus immersus and the Pyrenomycetes P. anserina (P. anM), P. comata (P. co), Nectaria crassa (N. cr) and Cephalosporium acromonium (C. ac). The Pyrenomycetes nad1-i4-orf1-containing strains, S. macrospora (S. mac), N. intermedia (N. inV), P. anserina (P. anS) and P. comata (P. co) form an independent cluster. The bacterial phylum is represented by Escherichia coli (E. co). Bacterial, plant, animal and fungal clusters with and without ORF1 are indicated by dotted lines.
The absence of the optional ORF. In the case of coxl, over 140 exonic nucleotides flanking the mono-orfic intron and 650 nucleotides in the intron, no sequence variation was detected when compared with the biorfic counterpart, except for the six substitutions already mentioned and an additional triplet (TTT) three nucleotides upstream from the end of the intron (see Figure 1B'). These results are in agreement with previous data showing that mitochondrial sequence divergence among the P. anserina strains is very low (CUMMINGS et al. 1990). They furthermore show that the differences between mono-orfic and biorfic introns containing genes are tightly clustered in close proximity to the optional sequences. The last six nucleotides of the nadl-i4 upstream exon and the first five nucleotides of the coxl-i7 downstream exon, which are known to be involved in base pairings and required for intron splicing (5' side of the P1 and 3' side of the P10 stems, respectively), remain identical in the mono- and biorfic organizations.

All the nucleotide differences observed between exons flanking biorfic and mono-orfic introns are silent even though, in several cases, positions other than the third in the codons differ. The nine nucleotide substitutions in nadl-i4 led to the formation of seven modified codons. According to the codon usage in the mitochondrial genes of P. anserina (CUMMINGS et al. 1990), even though unusual codons are found in both the exonic sequences (Figure 2, lines 4 and 5), the exonic sequence linked to the biorfic intron would be at least 1000 times less frequently used than the sequence linked to the mono-orfic (or to the equivalent sequence in the intronless locus). In the case of coxl-i7, four codons are modified when the optional ORF is absent, but there is no global significant difference in the probabilities of the exonic sequences (Figure 2, lines 10 and 11).

The nadl-i4 and coxl-i7 equivalent introns in other Podospora species: P. comata and P. curvicolllla: The organization of nadl-i4 and coxl-i7 equivalent introns was examined in six other wild-type strains of Podospora belonging to two different species. nadl-i4 equivalent intron in P. comata was found again either biorfic (P. co3) or mono-orfic (P. coT, P. co2), the upstream ORF (ORF1) being optional. In P. curvicolllla, the nadl-i4 equivalent intron was either mono-orfic (P. cuV, P. cuM) or absent (P. cuL). coxl-i7 equivalent intron was found exclusively in the mono-orfic form in the three P. comata strains and no such sequence was present in the three P. curvicolllla strains. The absence of equivalent nadl-i4 (in P. cuL) and coxl-i7 (in P. curvicolllla) mitochondrial introns, corresponds to a new (third) configuration for both genes (Figure 1, C and C'). As the nadl-i4-orfl and the coxl-i7-orf2 were found optional in the species P. anserina, the whole nadl-i4 and coxl-i7 introns are optional in the genus Podospora.

We have determined the nucleotide sequences of the exons flanking the nadl-i4 and coxl-i7 introns (or the corresponding regions in strains devoid of the intron) in the six additional Podospora strains. For nadl, the sequences are again of two types, with a high degree of similarity with the P. anserina equivalent exons flanking biorfic introns in the case of the P. co3 (biorfic) and with the P. anserina equivalent exons flanking mono-orfic introns in the case of the P. coT, P. co2, P. cuM, P. cuV (mono-orfic) and P. cuL (intronless) strains. In the case of the coxl-i7 downstream exon, all the P. comata mono-orfic and P. curvicolllla intronless strains present high degree of sequence similarity with the P. anserina equivalent exon bordering the mono-orfic introns. Alignment of the sequences with that of P. anserina strongly reinforces the previous correlation where six exonic consensus substitutions remain associated with the presence of the optional ORF (Figure 2, lines 3, 6, 7 and 12, 13).

Whereas the mitochondrial genome of Podospora strains, within the same species, is >99% identical, excepted the polymorphism for optional sequences, the exonic sequences adjacent to introns devoid of the optional ORFs (or of the whole intron) present much more similarity with their counterparts bordering mono-orfic intron in other species than with their counterparts bordering biorfic intron in the same species. Such comparisons were reported in Table 1 for the nadl-i4 locus (3'EX). The sequence comparison of the P. coT and P. co3 nadl-i4 mono- and biorfic introns revealed that they mainly differ by the presence of the nadl-i4-orf1 and associated exonic sequence: their catalytic core domain and ORF2 are very similar as is the case for P. anA and P. anM (Table 1). However, interspecific comparisons revealed two unexpected features. (1) The mean level of divergence found between the P. anserina and P. comata mono-orfic intron (~20%; see Table 1) is in contrast with the average of 5% found in several
TABLE 1
Pairwise comparisons in the *nadl-i* locus

<table>
<thead>
<tr>
<th></th>
<th>ORF1+</th>
<th>ORF1−</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S. ma6</td>
<td>P. anA</td>
</tr>
<tr>
<td>N. inV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>midEX</td>
<td>97</td>
<td>83</td>
</tr>
<tr>
<td>3' EX</td>
<td>87</td>
<td>90</td>
</tr>
<tr>
<td>ORF1</td>
<td>99</td>
<td>66</td>
</tr>
<tr>
<td>core</td>
<td>92</td>
<td>69</td>
</tr>
<tr>
<td>P. anA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>midEX</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>3' EX</td>
<td>87</td>
<td>94</td>
</tr>
<tr>
<td>ORF1</td>
<td>66</td>
<td>98</td>
</tr>
<tr>
<td>core</td>
<td>78</td>
<td>88</td>
</tr>
<tr>
<td>ORF2</td>
<td>75</td>
<td>72</td>
</tr>
<tr>
<td>P. co3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>midEX</td>
<td>81</td>
<td>100</td>
</tr>
<tr>
<td>3' EX</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>ORF1</td>
<td>67</td>
<td></td>
</tr>
<tr>
<td>core</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>ORF2</td>
<td>79</td>
<td></td>
</tr>
</tbody>
</table>

The comparisons (percentage of identity, in nucleotides) concern the 28 last nucleotides of *nadl-e4* (3'EX), the *nadl-i* intronic coding sequences (ORF1, ORF2), and the catalytic core (from the 5' side of the P1 to the 3' side of the alternative P10 stems) of the intron between various *Podospora* strains belonging to the same or different species or between strains of different genera. The alignment of *P. cot* and *P. anA* *nadl-i-orf2* nucleotide sequences (675 nucleotides) revealed 72% identity with never more than 13 matched adjacent nucleotides. One hundred fifty nucleotides in the middle part of the upstream exon (midEX) were used as a mean of evaluation of the degree of divergence between strains of different genera. *P. cuV* and *P. cot* *nadl-i* intron presents the same organization and shows more than 99% similarity in sequence; *P. cuV* was thus not included in the table.

Mobility of *nadl-i-orf1* and *cox1-orf2* in *Podospora*: Heteroplasmons, resulting from the vegetative fusion between mycelia of *Podospora* strains differing by the presence of the optional ORFs, were constructed. To avoid the incompatibility barrier and allow cytoplasm mixing, the donor and recipient strains were rendered isogenic for their nuclear genes (see MATERIALS AND METHODS). Strains containing the mitochondria of *P. anA* and *P. anS* were used as the ORF donors to recipient strains containing the mitochondria of *P. anM* (*Het anM/anA*), *P. anD* (*Het anD/anA*), *P. anV* (*Het anV/anA*), and *P. cot* (*Het cot/anS*). In each experiment, the transmission of several unlinked mitochondrial markers from the donor strains was monitored: in the heteroplasmon progeny, the percentage of homoplasmic
strains harboring a given marker (i.e., polymorphic restriction site) from the donor strain was used as a measure of the transmission rate (Figure 4). *nadl-i4* and *cox1-i7* were found in 100% of the homoplasmic strains whereas the polymorphic genetic markers from the donor were transmitted with an efficiency of <50% differing from one experiment to another (Table 2). Intronic ORFs thus appeared highly invasive in agreement with the behavior of mobile elements propagating via the homing mechanism described for group I introns (DuJON 1989). Moreover, the sequences of the new exon-intron junctions indicated that the transmission of the optional ORFs was, in all cases, accompanied by coconversion in the adjacent exon, bringing the exonic sequence characteristic of the optional ORF-containing strain (Figure 2, lines 4 and 10). Bidirectional coconversion was expected if the double strand break repair process, involved in group I intron homing, also applies to ORF mobility. Monitoring of the conversion inside the intron was not possible when the two parental strains contained identical intronic sequences (core and constitutive ORF). In the case of the *Het coT/anS* progeny, where the parental strains contained very divergent *nadl-i4* sequences (Table 1), the entire intron sequence (ORF1, core, ORF2) was found to be that of the donor strain (not shown).

**The *nadl-i4* and *cox1-i7* introns in other Pyrenomycete genera:** *Neurospora* and *Sordaria:* *nadl-i4* and *cox1-i7* equivalent introns, inserted at identical genomic sites (a highly conserved domain of the exonic amino acid sequence), have been described in mitochondrial DNA of Neurospora. Their structure is similar to that found in Podospora (see Figure 1). *nadl-i4, nadl-i4-orf1* and *nadl-i4-orf2* are optional in Neurospora as deduced from the three different organizations found in this genus (MOTA and COLLINS 1988; HAWSE et al. 1990).

The *nadl-i4* equivalent intron is mono- or bifurcated in some strains, containing either an equivalent ORF1 (*N. intermedia, strain Varkud*) or an equivalent ORF2 (*N. crassa, strain 74A*), or intronless in others (*N. crassa, strain Abbott 12*). Interestingly, the *nadl-i4* equivalent in *N. intermedia* was probably derived from a bieoric intron; it contains a full length and most likely functional ORF1 equivalent but a truncated ORF2 in which 47 codons of the *N. crassa* equivalent ORF were retained as well as the hypothetical alternate 3’-splice site that would ensure the translation of the free standing ORF before it was lost (Figure 1D). We have also undertaken the sequencing of the *nadl-i4* equivalent intron in a related fungus, *S. macrospora,* we found to present a bieoric organization with an exceptionally long and fragmentary ORF2 (C. H. SELLEM, unpublished results). The *cox1-i7* intron is optional in Neurospora. In the only strain where it has been described (*N. crassa Adiopodoume,* it presents the same mono- or bifurcated organization as in Podospora strains (FIELD et al. 1989). Alignment of the exonic sequences, adjacent to optional ORFs or in the homologous location in genes devoid of the intron, revealed that the correlation previously observed in Podospora is maintained throughout all the Pyrenomycete strains examined (Figure 2). Over a total of 29 strains examined, belonging to six species in three different genera, five consensus substituted positions in the *nadl-i4* equivalent upstream exon and four in the *cox1-i7* equivalent downstream exon are perfectly correlated with the presence of the corresponding optional ORF. The exonic sequences located upstream of the *nadl-i4-orf1* in various species and genera (*N. inV, P. anA, P. co3, S. ma6*) present much more similarity one to another than with the corresponding exonic sequences in strains of the same species devoid of ORF1 (*N. cv7, P. anM, P. coT*) that in turn appear to be more closely related to one another (Table 1).

**nadl-i4 region:** Sequence and comparison with more distantly related species: Among sequence banks no Pyreno-
mycete species other than those already mentioned harbor a nadl-i4 equivalent intron. Among filamentous fungi, Aspergillus nidulans (BROWN et al. 1983) and Asco- bolus immersus (C. H. SELLEM, unpublished results) are intronless (Figure 1C). A pairwise comparison of the sequences equivalent to the P. anserina nadl-i4 30 last nucleotides was undertaken for distantly related species. Even based on a very short sequence comparison, the resulting unrooted phylogenetic tree (inset in Figure 2) reflects rather well the evolution of the various kingdoms (bacteria, fungi, plants, animals) with only few cases of incorrect branching. The division of the fungal kingdom in two distant clusters, one of which included all the nadl-i4-orf1-containing strains and branching earlier in the tree, will be discussed below as the result of horizontal transfers of nadl-i4-orf1 and neighboring exonic sequences.

The correlation between the presence of the optional sequence and a consensus nucleotide sequence in the adjacent exon led us to evaluate the degree of similarity in optional against nonoptional ORFs in various wild-type strains of different genera. This has been done only for nadl-i4 since no biorfic coxl-i7 equivalent was found elsewhere than in P. anserina. The coxl-i7 equivalent introns found in sequence banks and not already mentioned belong to A. nidulans (WARING et al. 1984) and, surprisingly, to the liverwort Marchantia polymorpha (OHTA et al. 1993); in both cases they show a mono-orfic organization (Figure 1C'). Three of the four substitutions associated with the absence of the downstream ORF remain conserved. The opti- nal sequences are available for one strain of P. anserina and two species of Neurospora (MOTA and COLLINS 1988; HAWSE et al. 1990). They were compared with mono-orfic or biorfic nadl-i4 introns we have obtained for various Podospora strains from different species and for S. macrospora. Table 1 shows that, among the various strains examined, core and ORF2 evolved in a very similar way. The level of similarity observed within a species is high while it is much lower between species and genera. This is suggestive of the evolution of the mono-orfic intron from a common ancestor to the three genera. On the contrary, the nadl-i4-orf1 sequences of P. anserina and P. comata are very similar in otherwise divergent introns (core and ORF2) as are those of N. intermedia and S. macrospora suggesting, in those cases, horizontal transfer events had occurred.

The 17-bp sequence, located downstream from nadl-i4orf1 and upstream from the core structure in the P. anserina biorfic intron, was identified as an exact replica of the 3' end of the N.crassa nadl-i4 equivalent upstream exon (CUMMINGS et al. 1988). We found the same sequence in the 3' end of P. anserina nadl-i4 flanking mono-orfic introns (Figure 1A and B). Similarly, in P. comata, a sequence almost identical to the last 17 nucleotides of P. coT nadl-i4 is present at the equivalent location in the P. co3 biorfic intron. The six invariant last nucleotides of the nadl-i4, encoding the

---

TABLE 2

<table>
<thead>
<tr>
<th>Optional sequences</th>
<th>Polymorphic restriction sites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>coxl</td>
</tr>
<tr>
<td></td>
<td>Haell</td>
</tr>
<tr>
<td>Parental strains</td>
<td></td>
</tr>
<tr>
<td>Donor</td>
<td></td>
</tr>
<tr>
<td>P. anA</td>
<td>+</td>
</tr>
<tr>
<td>P. an$^*$</td>
<td>+</td>
</tr>
<tr>
<td>Recipient</td>
<td></td>
</tr>
<tr>
<td>P. anM</td>
<td>-</td>
</tr>
<tr>
<td>P. anD</td>
<td>-</td>
</tr>
<tr>
<td>P. coT$^*$</td>
<td>-</td>
</tr>
<tr>
<td>Heterokaryons</td>
<td></td>
</tr>
<tr>
<td>Het anM/anA</td>
<td>+ 18/18</td>
</tr>
<tr>
<td>Het coT/anS$^*$</td>
<td>+ 30/30*</td>
</tr>
</tbody>
</table>

Homoplasmic strains were issued from heterokaryons raised between (recipient/donor) isogenic strains bearing P. anA and P. anM (Het anM/anA), P. anA and P. anD (Het anD/anA) or P. coT and P. anS (Het coT/anS) mitochondria. The donor and recipient parental strains exhibit the A (or A’) and the B (or B’) organization, respectively (see Figure 1). The presence (+) of nadl-i4-orf1 and coxl-i7-orf2 was visualized on Southern blot of mitochondrial DNA digested with Haell as described in legend of Figure 4. The five independent mitochondrial genetic markers monitored in the progenies are polymorphic restriction sites: Haell digestions allowed the detection of the P. anserina (an) or P. comata (co) origin of the coxl-i1, nad5-i3 (see Figure 4) and cytB-i loci. In the same way, HaellI and HinfI restriction sites specific of the strain P. anA (A) in the nad5-i1 and coxl-i1 introns, respectively, were also monitored in the progenies. The number of homoplasmic strains (over the total examined) harboring the various markers from the donor strain, is indicated. The asterisks point out the characteristics of the parental strains and heterokaryon progeny (Het coT/anS) presented in Figure 4.
5' side of the P1 stem of the RNA structure, then constitute a direct repeated sequence flanking the optional ORF. A direct repeat organization was also found in the coxl-i7 biorfic intron: a 8-bp sequence overlapping the intron-exon junction is also present at the core-ORF2 junction. These repeats constitute the proximal alternative (upstream repeat) and the distal (downstream repeat) 3’-splice sites determined by the 3' side of the alternative and constitutive P9-0 and P10 stems flanking the last G of the intron (SELLEM and BELCOUR 1994). A comparable situation has been described for the Neurospora nadl-i4 (MOTA and COLLINS 1988), but direct repeats flanking both optional ORFs are not precise elements of pairing in the RNA secondary structure, as in Podospora. The significance of such short repeats (5–8 bp) flanking the optional ORF and of sequences (17 bp) found both in the exon bordering the mono-orfic and in the biorfic intron is not clearly understood, but they have been proposed to be involved in the acquisition or loss of the optional sequences (MOTA and COLLINS 1988; VADER et al. 1994).

**DISCUSSION**

The mitochondrial chromosome in wild-type strains of the filamentous Ascomycete P. anserina presents a previously unreported feature: the presence, within a species, of optional intronic ORFs. Two introns referred as nadl-i4 and coxl-i7 may present several organizations depending on the natural isolates examined: either biorfic (containing two ORFs) or mono-orfic (containing only one). The organization of these two introns in P. anserina offered the opportunity to investigate the mobility of the optional ORFs as independent genetic elements. Furthermore, the systematic clustered nucleotidic differences we found between exons flanking mono- and biorfic introns led us to question the evolutionary origin of these ORFs and the mechanisms involved in their transfer.

**Arrival of intronic ORFs in the nadl-i4 and coxl-i7 introns of Podospora:** The large number of clustered substitutions we found in P. anserina mitochondrial exons adjacent to the optional ORFs in biorfic nadl-i4 and coxl-i7 introns, relative to their mono-orfic counterparts, is clearly unusual since much more than 99% conservation of sequence has been found in the overall genome of the wild-type strains belonging to this species (CUMMINGS et al. 1990). The correlation established between the presence of the optional ORFs and the nucleotidic sequence in the adjacent exon is strong enough to be extended to all other Pyrenomycete species we have examined. What does this correlation mean? The presence of an optional sequence in a species raises the question of its loss or gain in the course of evolution. When the sequence is an intron, optional- ity is generally explained by a steady state between loss and gain (DUJON 1989). How to loose and gain group I introns seems now to have been elucidated: loss proceeds via reverse transcription of mature RNAs (GARGOURI et al. 1983; LEVRA-JUILLET et al. 1989; SAINSSARD-ChANET et al. 1993) and gain (homing), via double-strand breaks and conversion of the intronless allele using the intron containing gene as a template (DUJON 1989; LAMBOWITZ and BELFORT 1993).

The known mechanism of loss could not be invoked for the nadl-i4 and coxl-i7 optional ORFs, since no splicing event generating RNA molecules cleanly devoid of the optional sequences is expected or detected. Anyhow, all the spliced RNA molecules issued from nadl and coxl transcripts containing the optional ORF exhibited, as expected, the exon characteristic of the biorfic intron-containing strains (SELLEM and BELCOUR 1994). A loop out at the DNA level, between direct repeats flanking the optional ORFs, could exceptionally promote their deletion as it has probably been the case for part of the nadl-i4-orf2 in a close ancestor of N. intermedia (MOTA and COLLINS 1988). According to the presence of similar direct repeats, complete excision of the two optional ORFs would also be possible in Podospora nadl-i4 and coxl-i7 intron. However, such an ORF loss would not be accompanied by concerted modifications in the region of excision. Loss of the intronic ORF, unless invoking a new mechanism with coconversion, could then not be invoked in a steady-state scenario where optional sequences are periodically lost and gained. Such a mechanism of loss, first invoked by HENSCENS and coworkers (1983) for an intron, implying cut and repair using a foreign ORF-less genome as template, cannot however be completely excluded. In contrast, gain of intronic ORFs could be explained by a mechanism equivalent to that proposed for group I intron homing that perfectly accounts for the modifications encountered in the sequence flanking the insertion region. Moreover, in the case of nadl-i4 and coxl-i7 optional ORFs, homing (gain) is consistent with their observed invasiveness.

**Mobility of intronic ORFs inside the genus Podospora:** In the progeny of heteroplasmons constructed between biorfic and mono-orfic intron-containing strains, we have shown that the nadl-i4-orf1 and the coxl-i7-orf2 appeared highly invasive. Indeed, all the homoplasmons (100%) tested had acquired the optional ORF from the biorfic donor strains. Moreover, in all the cases, the transfer of the optional ORFs was systematically accompanied by the transfer of the associated exonic-specific sequence. Searching for an expected but limited coconversion inside the intronic sequence in the only case where this sequence was different between the parental strains (nadl-i4 in Het coT/anS), revealed that the whole intron of the donor strain has been transferred in the progeny. However, it could not be definitively concluded that the mobile entity was the entire intron. Indeed, the very high divergence in core and ORF2 sequences between the donor and the recipi-
ent strains (Table 1) could be responsible for an inevitable coconversion of the intron initiated by the transfer of the ORF as the mobile element. The alternative and less likely interpretation would be the transfer of the intrinsic ORF together with the whole intron, proceeding via an intron replacement; the biorfic intron taking the place of the mono-orfic one. These experiments nevertheless constitute the first examples of driven ORF transfer from an ORF-containing intron to the cognate ORF-less allele, probably miming the horizontal events that have recently taken place in the nad1-i4 and cox1-i7 introns of Podospora.

The acquisition of nad1-i4-orf1 and cox1-i7-orf2 is recent in Podospora: The pattern of presence/absence of optional ORFs suggests that the steady state has not been reached and that they are currently invading the population. While for an intron steady state would mean periodical losses and gains, in the case of mobile intronic ORFs, steady state would be all the introns being invaded. The acquisition of nad1-i4-orf1 and cox1-i7-orf2 by Podospora may be relatively recent, so that insufficient time has elapsed to permit evolution of the overall genome in two strains of the same species differing by the presence of the optional ORF. The presence of a 17-bp sequence, both in the 3' noncoding region of the optional nad1-i4-orf1 in biorfic intron (P. anS and P. co3) and in the 3' end of the nad1-e4 exon adjacent to mono-orfic intron, is surprising. This could suggest that the optional ORF has integrated into the exonic sequence adjacent to mono-orfic intron, bringing at the same time a sequence equivalent (but sufficiently modified) to the one displaced by the insertion. An analogous situation was already described (Paquin et al. 1994). In the case of P. anserina, the identity in the 17-bp sequence, present in mono- and biorfic loci, could be an additional argument for the recent acquisition. Data given in Table 1 agree with the current acquisition of nad1-i4-orf1 and associated exonic modifications by the Podospora nad1-i4 mono-orfic intron (core and ORF2). P. anserina (P. anA) and P. comata (P. co3) nad1-i4-orf1 present much more similarity one with another than the overall intron does. This strong similarity at both amino acid and nucleotide levels implies not only similar functional constraints but also a recent common origin.

Origin of the optional ORFs: The acquisition of additional intronic ORFs by Podospora nad1-i4 and cox1-i7 introns, together with modifications in the flanking exons, are most probably the result of recent horizontal transfers via homing-like events from other fungal nad1 and cox1 mitochondrial genes. Indeed, searching in sequence banks has revealed that the exonic sequence adjacent to the studied optional ORFs present no significant similarity with any other gene than the mitochondrial nad1 and cox1. In addition, the codon usage of the two ORFs appears to conform to that used in fungal mitochondria, whereas in the case of nad1, that of the adjacent exon is especially abnormal. This could mean that the two optional ORFs were present in mitochondrial genomes for a sufficiently long time to have adapted their codon usage, whereas the associated exonic sequences are probably assigned to different constraints preventing such adaptation. The especially highly improbable codon usage in exons flanking biorfic nad1-i4 intron led us to search, in sequence banks, for an hypothetical nad1 foreign donor gene. A phylogenetic tree was constructed from pairwise comparisons of equivalent nad1-e4 50 last nucleotides among various unrelated species including those containing mono-orfic and biorfic nad1-i4 introns. Although based on a very short sequence, this tree evidences the horizontal transfer of nad1-i4-orf1 in various fungal strains (Figure 2, inset) by the independent clustering of all the ORF1-containing strains. The branch point could give an idea of the possible foreign donor organism that appears to be more distantly related to Podospora (Pyrenomycete) than Podospora is to Aspergillus (Plectomycete).

Horizontal transmission of introns between different organisms, although never observed in a laboratory, has previously been invoked to explain the presence of related introns in otherwise more divergent genes (Lang 1984; Waring et al. 1984; Trinckle and Wolf 1986) or in different locations (Michel and Cummings 1985; Colleaux et al. 1990). Such lateral transmission events were also proposed for the ORFs containing the GY-YIG motif (found in Podospora and Neurospora mitochondrial group I introns), as a genetic exchange between a close ancestor of the Neurospora-Podospora group and the bacteriophage T4 genome (Michel and DuJon 1986).

Evolution of the nad1-i4 intron: In Figure 5 we have schematically drawn, from data of different sources taking into account both biological (see MATERIALS AND METHODS) and sequence data, the lineage of the various strains studied in this work. Clearly, numerous events of gain and loss of optional sequences are needed to explain the current status of the various nad1-i4 equivalent introns examined. We have indicated the most probable events that also take into account the sequence similarity found over the three domains of the intron between the strains, species and genera of the Pyrenomycetes examined (Table 1). When very similar sequences were found in divergent contexts, we have invoked horizontal transfer events: between P. anA and P. co3 and between N. intermedia and S. macrospora for nad1-i4-orf1 and nearby exonic sequences, and between P. comata and P. eurycolla for the mono-orfic nad1-i4 intron. The rather poor degree of sequence similarity found between Podospora and Neurospora nad1-i4-orf1 led us to propose independent horizontal transfer events from ORF1 donors having evolved separately. Indeed, a horizontal transfer between Neurospora and Podospora would suppose that the intronic optional ORFs (66% similarity) evolve much more rapidly than
the overall mitochondrial genome (~80% similarity). The transfer of *nad1*-i4 mono-orfic intron from *P. curvicolla* to *P. comata* implies a prior loss of the intron in *P. comata* and explains why the degree of similarity of the *nad1*-i4 introns (core and ORF2) is low between two closely related species (*P. anserina* and *P. comata*) and is high between two more distant ones (*P. comata* and *P. curvicolla*). All the transfer events proposed on the basis of sequence similarities (Table 1) now agree with the presence of an ancestral *nad1*-i4 mono-orfic intron that predates in the three Pyrenomycetes genera. Figure 5 gives a parsimonious scenario that also takes into account the loss of intronic sequences such as *nad1*-i4-orf2 in *N. inV* (loop out of the ORF) and occasional loss of the mono-orfic intron (*P. comata*, *P. cul* and *N. crA*). It could be predicted that loss of biorfic introns may also occur: according to the known mechanism of loss, this would result in an intronless gene bearing an upstream exon of the biorfic type. No such strains have yet been found among the 23 strains examined in *Podospora*, neither in the case of *nad1*-i4 nor *cox1*-i7, suggesting again that the acquisition of the two optional ORFs and associated exonic sequences is very recent.

Our results greatly support the idea that group I intronic ORFs are mobile by themselves and have found a safe location in preexisting introns. The case of biorfic introns is interesting since the acquisition of the optional ORFs is recent and provides evidence that the mechanism of their mobility involves modifications of adjacent exons, thus playing a role in the modular evolution of the mitochondrial genomes.

We are grateful to E. DuFour for technical help in the preparation of the figures and to S. Haddad for sequencing assistance. This work was supported by grants from the GREG (no. 71/94) and from the Fondation pour la Recherche Médicale.

**LITERATURE CITED**


Lang, B. F., 1984 The mitochondrial genome of the fission yeast Schizosaccharomyces pombe: highly homologous introns are inserted at the same position of the otherwise less conserved codl genes in Schizosaccharomyces pombe and Aspergillus nidulans. EMBO J. 3: 1299–1313.


Communicating editor: M. R. Hanson