Evolutionary Relationships among Putative RNA-Dependent RNA Polymerases Encoded by a Mitochondrial Virus-like RNA in the Dutch Elm Disease Fungus, *Ophiostoma novo-ulmi*, by Other Viruses and Virus-like RNAs and by the Arabidopsis Mitochondrial Genome

Yiguo Hong,* Thomas E. Cole,* Clive M. Brasier,† and Kenneth W. Buck*†

*Department of Biology, Imperial College of Science, Technology and Medicine, Prince Consort Road, London SW7 2BB, United Kingdom; and †Forest Research Station, Alice Holt Lodge, Farnham, Surrey, GU10 4LH, United Kingdom

Received February 23, 1998; returned to author for revision March 16, 1998; accepted April 7, 1998

The nucleotide sequence (2617 nucleotides) of virus-like double-stranded (ds) RNA 3a in a diseased isolate, Log1/3-8d² (Ld), of the ascomycete fungus *Ophiostoma novo-ulmi* has been determined. One strand of the dsRNA contains an open reading frame (ORF) with the potential to encode a protein of 718 amino acids, and the complementary strand contains two smaller ORFs with the potential to encode proteins of 178 and 182 amino acids, respectively. The large ORF contains 12 UGA codons which code for tryptophan in ascomycete mitochondria and has a codon bias typical of mitochondrial genes, consistent with the localization of Ld dsRNAs within the mitochondria. The amino acid sequence contains motifs characteristic of RNA-dependent RNA polymerases (RdRps). This putative RdRp was shown to be related to putative RdRps of mitochondrial dsRNAs of another ascomycete and a basidiomycete fungus and also to a putative RdRp encoded by the mitochondrial genome of *Arabidopsis thaliana*. In multiple sequence alignments, the fungal mitochondrial dsRNA-encoded RdRp-like proteins formed a cluster, ancestrally related to the RdRps of the yeast 20S and 23S RNA replicons and of the positive-stranded RNA bacteriophages of the Leviviridae family, but distinct from RdRps of other families and genera of fungal RNA viruses and related plant and animal RNA viruses. Northern blot analysis with RNA 3a strand-specific probes indicated that nucleic acid extracts of Ld contain more single-stranded (positive-stranded) RNA than dsRNA, consistent with an evolutionary relationship between RNA 3a and positive-stranded RNA phages.© 1998 Academic Press

Key Words: *Ophiostoma novo-ulmi* mitochondrial virus-like RNA; RNA-dependent RNA polymerase; *Arabidopsis thaliana* mitochondrial genome; evolution of mitochondrial virus-like RNAs.

INTRODUCTION

A diseased isolate, Log1/3-8d² (hereafter designated Ld), of the Dutch elm disease fungus, *Ophiostoma novo-ulmi*, was reported to contain 10 virus-like double-stranded (ds) RNA components with approximate sizes (kbp) of 3.49 (RNA-1), 3.03 (RNA-2), 2.69 (RNA-3), 2.43 (RNA-4), 2.33 (RNA-5), 2.21 (RNA-6), 0.95 (RNA-7), 0.92 (RNA-8), 0.48 (RNA-9), and 0.33 (RNA-10) (Rogers et al., 1986). Analysis of single-conidal and yeast-phase isolates derived from Ld showed that loss of dsRNAs 4, 7, and 10 results in reversion to the healthy phenotype (Rogers et al., 1986). The Ld dsRNAs were found in purified mitochondrial preparations, in which they were resistant to ribonuclease digestion, suggesting a mitochondrial location (Rogers et al., 1987). Furthermore, isolate Ld had a greatly reduced level of mitochondrial cytochrome oxidase, consistent with a mitochondrial target for one or more of dsRNAs 4, 7, and 10 (Rogers et al., 1987). The Ld dsRNAs, and virus-like dsRNAs associated with other diseased phenotypes of *Ophiostoma novo-ulmi*, have potential for the development of biological control agents for Dutch elm disease (Brasier, 1986; Sutherland & Brasier, 1997).

Recently the precise sizes of dsRNA 7 (1057 bp) and dsRNA 10 (317±30 bp) were determined by nucleotide sequence analysis. Neither of these RNAs could encode proteins of more than 41 amino acids and therefore both presumably require one of the other dsRNAs for their replication. DsRNA 10 was found to be a defective-like RNA, consisting of a mosaic of sequences derived from dsRNA 7 (Hong et al., 1998). We have also recently shown, using a higher resolution polyacrylamide gel electrophoresis (PAGE) system, that dsRNAs 1 and 3 can each be resolved into two components, designated 1a, 1b, 3a, and 3b, respectively (T. E. Cole, B. Müller, Y. Hong, C. M. Brasier, and K. W. Buck, unpublished results). To obtain more information about the functions of the dsRNA components of Ld, we have determined the complete nucleotide sequence of dsRNA 3a and show that it has the potential to encode an RNA-dependent RNA polymerase-like protein. We have also shown that RNA 3a exists in a predominantly single-stranded (ss) (positive-stranded) form and have analyzed evolutionary relationships between putative RdRps encoded by RNA 3a.
and those encoded by other fungal viruses and virus-like RNAs by animal, bacterial, and plant RNA viruses and by an Arabidopsis mitochondrial DNA gene.

RESULTS AND DISCUSSION

Production of cDNA clones and nucleotide sequence of dsRNA 3a

A library of cDNA clones was obtained from gel-purified dsRNA. To prove that the clones obtained were unique to dsRNA 3a and had not arisen from contamination from other RNAs, dsRNAs from isolates Ld and W2tol1, sci59, sci60, and sci64, which contained subsets of the Ld dsRNA components, were analyzed by PAGE, followed by Northern blotting and hybridization with riboprobes synthesized from the cDNA clones. An example is shown in Fig. 1B. A single dsRNA with the same electrophoretic mobility of dsRNA 3a, present only in Ld and sci59, hybridized with the cDNA clone riboprobe. No hybridization was obtained with dsRNA from W2tol1, sci60, and sci64, the components (1a to 6) of which are tabulated in Fig. 1A. The dsRNA components of sci59 and 64 differ only in that sci59 contains dsRNA 3a and sci64 lacks it. The positive hybridization of the probe with dsRNA from sci59 and the absence of hybridization with sci64 therefore unequivocally identifies dsRNA 3a as the source of the cDNA clone. The absence of hybridization with dsRNA from sci60 and W2tol1, and the presence of a single hybridizing band with Ld dsRNA, is consistent with this assignment. Because of the prolonged electrophoresis needed to separate dsRNA 3a and dsRNA 3b, dsRNAs 7 to 10 had run off the end of the gel. Further Northern blots of agarose gels of dsRNA of Ld and derived single-conidial isolates, electrophoresed for shorter periods so that dsRNAs 7 to 10 were retained on the gel, showed that the dsRNA 3a cDNA probes did not hybridize with dsRNAs 7, 8, 9, or 10 (not shown).

The overlapping cDNA clones used to obtain the complete sequence of dsRNA 3a (2617 bp) are illustrated in Fig. 2A.

A long open reading frame in the (+) strand of dsRNA 3a encodes an RNA-dependent RNA polymerase-like protein

The nucleotide sequence of dsRNA 3a was examined for the presence of open reading frames (ORFs) on both strands. As Ld dsRNAs are located within the O. novo-ulmi mitochondria and O. novo-ulmi is an ascomycete fungus, the sequence UGA, which acts as a terminator for translation by cytoplasmic ribosomes, was taken to encode tryptophan as in the mitochondria of most other ascomycete fungi (Osawa et al., 1992; Paquin et al., 1997). A UGA codon specifying tryptophan has previously been identified in a mitochondrial intron-encoded protein of O. novo-ulmi (Abu-Amero et al., 1995). An ORF, termed V1, with the potential to encode a protein of 718 amino acids was found to extend almost the full length of the (+)-strand RNA (Fig. 2B). This ORF contains 12 UGA codons fairly evenly spaced. Analysis of this protein revealed several amino acid sequence motifs typical of RNA-dependent RNA polymerases (RdRps) (Fig. 3). Motifs A, B, C, and D correspond to those described by Poch et al. (1989); motifs similar to motifs A, B, and C have also been described by Habili and Symons (1989), Bruenn (1991) and Koonin (1991). The motif designated E is similar to motif 7 of Bruenn (1993) and part of motif VII of Koonin (1991). Nucleotide polymorphisms were detected at 12 positions in the RNA 3a sequence. Only five of these would result in amino acid changes in the V1 protein (V221I, C568R, D577H, T580A, L663S), none of which were within the conserved RdRp motifs.

Previously an ascospore isolate derived from O. novo-ulmi Ld was shown to contain only dsRNA 2 and it was suggested that this RNA encodes an RdRp (Rogers et al., 1986, 1988). The finding that RNA 3a encodes an RdRp-like protein, and does not hybridize with dsRNA 2 or any of the other Ld dsRNAs, indicates that at least two of the Ld dsRNAs encode RdRps.
FIG. 2. (A) Maps of cDNA and RACE clones used to construct the nucleotide sequence of RNA 3a. The complementary sequences corresponding to nucleotides 423±397 and 564±539 were used as the primers for obtaining RACE clones for the 5' end of the (+) strand of RNA 3a. The sequences corresponding to nucleotides 1747±1570 and 1670±1694 were used as the primers for obtaining RACE clones for the 5' end of the (−) strand, complementary to the 3' end of the (+) strand of RNA 3a. All the clones shown were obtained from independent cDNA reactions and were sequenced in both directions. The RNA 3a sequence is available from the EMBL, GenBank, and DDBJ databases with the Accession No. AJ004930. (B) ORFs in the (+) and (−) strands of dsRNA 3a, assuming UGA codons specify tryptophan. The shaded region in ORF V1, designated POL, contains the RdRp motifs A to E. (C) Potential secondary structures of (i) the 5' terminal 20 nucleotides and (ii) the 3' terminal 116 nucleotides of the (+) strand of RNA 3a. The RNA was folded with the program MFOLD and in (ii) the result was displayed graphically with the program SQUIGGLES. All nucleotide numbers in A–C are for the positive strand.
Possible other functional domains in the dsRNA 3a RdRp-like protein

RNA replication frequently requires a helicase in addition to an RdRp. A helicase can be encoded on a separate RNA or can be a domain of a multifunctional protein that also contains an RdRp domain (Buck, 1996). A search was therefore made in the protein encoded by the V1 ORF for amino acid motifs typical of RNA helicases. Although no complete identity with any of the helicase consensus motifs (Koonin and Dolja, 1993) was obtained, a sequence AamrGKL, encoded by nucleotides 1487–1510, differed from the ATP-binding consensus A/GxxxxGKS/T only in the last amino acid. However, the position of this sequence between RdRp motifs B and C suggests that it is unlikely to correspond to part of a helicase domain, as helicase domains are usually separate from the polymerase domains (Buck, 1996). The absence of typical helicase domains would not exclude

---

**FIG. 3.** Alignment of amino acid sequences of RdRp-like proteins encoded by mitochondrial dsRNAs from *O. novo-ulmi* (Onu) (this study), *C. parasitica* (Cp) (Polashock and Hillman, 1994), *R. solani* 1A1 (Rs) (EMBL/GenBank/DDBJ Accession No. U51331), and by the *A. thaliana* mitochondrial genome (At) (Unseld et al., 1997). The alignment was made using the CLUSTAL W program. Conserved motifs and regions (A to E, X, Y) are underlined. Amino acid sequences which are identical in two or more sequences are shown in bold. At positions where there are two pairs of identical amino acids, one pair is shown in bold italics. Asterisks above the sequences indicate identical amino acids in all four sequences. Asterisks below the sequences indicate identical amino acids in the three fungal dsRNA-encoded sequences.

---

Obtained, a sequence AamrGKL, encoded by nucleotides 1487-1510, differed from the ATP-binding consensus A/GxxxGKS/T only in the last amino acid. Similar deviations from the consensus have been described for known ATP-binding proteins (Koonin, 1993). However, the position of this sequence between RdRp motifs B and C suggests that it is unlikely to correspond to part of a helicase domain, as helicase domains are usually separate from the polymerase domains (Buck, 1996). The absence of typical helicase domains would not exclude
helicase activity, since blue tongue virus VP6 protein has been shown to bind ATP and exhibit RNA-dependent ATPase and helicase activities in the absence of typical helicase motifs (Stäuber et al., 1997).

In addition, a putative eukaryotic and viral aspartic protease active site "AMADs/TQsVmA" was also identified within the conserved region X of the dsRNA-3a-encoded RdRp-like protein (Fig. 3), which has only a single amino acid change (underlined) compared with the consensus motif AMADS/TGTQxVxA. However, this is unlikely to be part of a functional protease domain as it lacks the G residue which is invariant in both cellular and viral aspartic proteases (Davies, 1990; Dougherty and Semler, 1993).

Detection of the viral single-stranded (positive-stranded) form of RNA 3a

ssRNA (2 M LiCl fraction) and dsRNA (5 M LiCl fraction) from O. novo-ulmi isolate Ld, before and after treatment with S1 nuclease and/or DNase I, were electrophoresed in an agarose gel, denatured with alkali, blotted onto a membrane, and hybridized with strand-specific riboprobes, synthesized from a dsRNA 3a cDNA clone. A band with the mobility of dsRNA 3a, which was resistant to both DNase I and S1 nuclease, was detected by both (+)-strand-specific (Fig. 4A, lanes 1 to 3) and (−)-strand-specific (Fig. 4B, lanes 1 to 3) probes, confirming that it is double stranded. The ssRNA fraction also gave a single band with the (+)-strand-specific probe, which was resistant to DNase I, but susceptible to degradation by S1 nuclease (Fig. 4A, lanes 4 to 7), confirming its single-stranded nature. In this gel system, the ssRNA and dsRNA bands had fortuitously comigrated. However, the size of the ssRNA band was determined by coelectrophoresis with ssRNA markers derived by in vitro transcription of cDNA clones of RNA 3a in the vector LITMUS 28 to be 2.6 kb (not shown), the size expected for the full-length positive strand of dsRNA 3a. The relative intensities of the (+)-strand ssRNA bands exceed those of the dsRNA bands significantly (Fig. 4A, compare lanes 4 to 6 with lanes 1 to 3). The amounts of ssRNA and dsRNA analyzed were derived from the same amounts of total nucleic acids. Hence, either there is much more ssRNA 3a than dsRNA 3a in isolate Ld or the efficiencies of transfer of dsRNA 3a to the membrane, and/or hybridization with the probe, is much less efficient than those of ssRNA 3a. The former explanation seems more likely, since the RNA was denatured and partially fragmented with alkali before blotting, so all the RNA samples should have been in the ss form prior to transfer and hybridization. No band was detected in the ssRNA fraction with the (−)-strand-specific probe (Fig. 4B, lanes 4 to 6).

5' and 3'-untranslated sequences in the RNA 3a (+) strand

There are 268 nucleotides prior to the start of ORF for the RdRp-like protein in the RNA 3a (+) strand and 192 nucleotides downstream of the UAA termination codon of this ORF. Since 5' and 3'-untranslated sequences are likely to contain promoters for RNA replication, and can often be folded into specific secondary structures (Buck, 1996), it was of interest to examine the RNA 3a (+)-strand leader and trailer sequences for potential secondary structures. Using the program MFOLD, it was shown that the 5'-terminal 20 nucleotides can be folded into a stem-loop structure (ΔG = −7.9 kcal/mol) (Fig. 2Ci). This structure was maintained when the entire 5'-untranslated sequence (nucleotides 1±268) was folded to give the potentially most stable structure (ΔG = −53.5 kcal/mol) and the six next potentially most stable structures (ΔG = −531 to −515 kcal/mol). The 3'-terminal 116 nucleotides could be folded to give a potentially
stable, although imperfect, stem-loop structure ($\Delta G = -21.5$ kcal/mol) (Fig. 2Cii). This was also maintained when the entire 3'-untranslated region (nucleotides 2426a2617) was folded.

The ability to form stem-loop structures at the 3' and 5' termini is similar to that described for RNA 7 and RNA 10 (Hong et al., 1998), except that there was no significant complementarity between the 3' and 5' termini, so that RNA 3a could not be folded into a potentially stable panhandle structure as described for RNAs 7 and 10.

There was no significant sequence similarity between the 3'- and 5'-terminal sequences of the (+) strand of RNA 3a and the 3'- and 5'-terminal sequences of RNA 7 and RNA 10 (Hong et al., 1998). The replication of RNA 7 and RNA 10 therefore probably depends on one of the other nine Ld dsRNAs.

Open reading frames in the (−) strand of dsRNA 3a

ORFs, designated C1 and C2, with the potential to encode proteins of 182 and 178 amino acids, respectively, were found on the complementary, (−) strand of dsRNA 3a (Fig. 2B). ORFs C1 and C2 contained two and three UGA codons, respectively. There was no significant amino acid sequence similarity between the putative proteins encoded by ORFs C1 and C2 and any protein sequence in the OWL database. The failure to detect either full-length or subgenomic (−) strand ssRNA of RNA 3a (Fig. 4B, lanes 4 to 6) suggests that the occurrence of ORFs C1 and C2 may be fortuitous and that they are probably not expressed.

The RdRp-like protein of O. novo-ulmi Ld RNA 3a is related to RdRp-like proteins of both ascomycete and basidiomycete fungi

The RdRp-like protein encoded by O. novo-ulmi Ld RNA 3a was found to be related to an RdRp-like protein encoded by a 2.7-kbp mitochondrial dsRNA from another ascomycete fungus, Cryptocentria parasitica NB631 (Polashock and Hillman, 1994; Polashock et al., 1997), particularly in the central region containing the conserved RdRp amino acid sequence motifs (Table 1; Fig. 3). It is noteworthy that the ORF encoding the C. parasitica RdRp-like protein also contained multiple UGA codons (Polashock and Hillman, 1994). Analysis of the codon usage by the O. novo-ulmi and C. parasitica RdRp ORFs indicated that 68.5 and 69.1% of the codons, respectively, have either A or U in the third position. Such codon preferences are typical of fungal and plant mitochondrial genes and may, at least partially, reflect the A+U-rich nature of mitochondrial genomes (Cummings et al., 1990; Paquin and Lang, 1996; Unseld et al., 1997). Both the O. novo-ulmi and C. parasitica mitochondrial dsRNAs have high proportions of A + U residues (61.9 and 63.4%, respectively). It is also noteworthy that the ORFs encoding the RdRp-like proteins terminated in UAA on both dsRNAs; UAA is the preferred, or sometimes the only, stop codon used in fungal mitochondria (Paquin et al., 1997). The RdRp-like ORF in the C. parasitica dsRNA contains a UAG codon 15 codons downstream of the proposed AUG initiation codon (Polashock and Hillman, 1994). Some chytridiomycete fungi use UAG as a codon for leucine in their mitochondria, although this has not been observed previously in ascomycete fungi (Paquin et al., 1997).

A 3.6-kbp dsRNA which partially copurifies with mitochondria of Rhizoctonia solani, the anamorph (asexual form) of the basidiomycete fungus Thanatephorus cucumeris (Jian et al., 1997), has been reported to encode an RdRp-like protein, related to the C. parasitica RdRp-like protein (Lakshman and Tavantzis, 1995), but no detailed analysis of the sequence has been published. Examination of the R. solani 1A1 3.6-kbp dsRNA (Database Accession No. U51331) for ORFs revealed that it could encode an RdRp-like protein of 754 amino acids, provided that a frameshift occurred somewhere between nucleotides 1555 and 1662. Sequence comparisons (Fig. 3; Table 1) showed that this protein was related to the RdRp-like proteins encoded by both O. novo-ulmi Ld RNA 3a and C. parasitica NB631 dsRNA and suggests that the frameshift in the R. solani RdRp-like sequence is a (−1) frameshift at nucleotide 1561 (shift between glycine at position 380, encoded by reading frame 2, and threonine at position 381, encoded by reading frame 1). There was a high degree of sequence conservation between the three RdRp-like proteins not only in motifs A to E, but also in two newly identified regions designated X and Y (Fig. 3). Further analysis of the R. solani 3.6-kbp dsRNA showed that it has 56.8% A + U residues with a preference for codons ending in A or U (61.3% of the codons for the RdRp-like protein end in A or U), although both these

<table>
<thead>
<tr>
<th></th>
<th>Onu</th>
<th>Cp</th>
<th>Rs</th>
<th>At</th>
</tr>
</thead>
<tbody>
<tr>
<td>Onu</td>
<td>D</td>
<td>22.8</td>
<td>26.2</td>
<td>12.3</td>
</tr>
<tr>
<td>Cp</td>
<td>61.8</td>
<td>D</td>
<td>20.4</td>
<td>10.6</td>
</tr>
<tr>
<td>Rs</td>
<td>64.0</td>
<td>60.8</td>
<td>D</td>
<td>10.2</td>
</tr>
<tr>
<td>At</td>
<td>56.2</td>
<td>51.7</td>
<td>46.1</td>
<td>D</td>
</tr>
</tbody>
</table>

Note. Amino acid sequence identities between pairs of sequences were determined using the GAP program. Top right triangle: amino acid sequence identities between complete RdRp-like protein sequences. Absent amino acids in gaps, including terminal gaps, were treated as mismatches. This introduces a substantial penalty for the much smaller size of the A. thaliana RpRp-like protein. Lower left triangle: amino acid sequence identities between sequences of motifs A, B, C, D, and E. There were no gaps in these sequences. RdRp-like proteins were encoded by O. novo-ulmi Ld dsRNA-3a (Onu); C. parasitica NB631 dsRNA (Cp); R. solani 1A1 3.6-kbp dsRNA (Rs); A. thaliana mitochondrial ORF251 (At).
properties were less marked than with the O. novo-ulmi Ld and C. parasitica NB631 mitochondrial dsRNAs.

Fungal viruses and virus-like RNAs are transmitted by intracellular routes, involving somatic or sexual cell fusions, generally limiting virus transmission to genotypes of the same or closely related species (Buck, 1986; Ghabrial, 1994). As a result, fungal viruses are believed to have coevolved with their hosts (Buck, 1998). The possibility therefore exists that mitochondrial dsRNA elements were present in fungi before the divergence of ascomycetes and basidiomycetes. Mitochondrial virus-like RNAs have been described in other basidiomycete fungi, such as Puccinia spp. (Kim and Klassen, 1989), and may be widespread. The finding that the RdRp-like protein encoded by O. novo-ulmi Ld dsRNA 3a is slightly more closely related to that encoded by the R. solani dsRNA than to that encoded by the C. parasitica dsRNA (Table 1) is surprising. This suggests differing rates of evolution of the dsRNAs within their respective ascomycete and basidiomycete hosts. An alternative explanation, that of rare horizontal transmission between these ascomycete and basidiomycete hosts, seems less likely in view of important differences between the R. solani dsRNA sequence and the O. novo-ulmi and C. parasitica dsRNA sequences. For example, the R. solani dsRNA contains no UGA codons in the sequence encoding the RdRp-like protein, all the 14 tryptophan residues (Fig. 4) being specified by UGG codons, whereas UGA is the preferred codon for tryptophan in both the O. novo-ulmi and C. parasitica dsRNAs. It is not known if proteins encoded by the R. solani mitochondrial genome contain tryptophan residues specified by UGA codons. Mitochondrial UGA codons are believed to have arisen independently in ascomycetes and basidiomycetes (Paquin et al., 1997). The only basidiomycete fungus for which mitochondrial DNA sequence data are available, Schyzophyllum commune, does use UGA codons to specify tryptophan residues in the mitochondria, but not in all of its mitochondrial proteins (Phelps et al., 1988; Paquin et al., 1995). Any possible horizontal transmission of virus-like RNA is unlikely to have occurred recently, because, outside the conserved RdRp motifs, the sequence divergence between the three RdRp-like proteins is high, with only 20±26% amino acid sequence identity over the whole proteins (Table 1).

Relationship of the O. novo-ulmi Ld RNA 3a RdRp-like protein to a putative RdRp-like protein encoded by the Arabidopsis mitochondrial DNA genome

The RdRp-like protein encoded by O. novo-ulmi Ld dsRNA 3a was also found to be related to an RdRp-like protein encoded by an ORF in the A. thaliana mitochondrial DNA genome (Table 1). Relationships between the A. thaliana RdRp-like protein and those encoded by C. parasitica NB631 and R. solani 1A1 dsRNAs have also recently been reported (Marienfeld et al., 1997). A multiple alignment of the amino acid sequences of the RdRp-like proteins encoded by O. novo-ulmi Ld dsRNA 3a, the C. parasitica NB631 and R. solani 1A1 dsRNAs, and the A. thaliana mitochondrial ORF251 is shown in Fig. 3. There are 45 positions in which amino acids in all four sequences are identical. These are mainly clustered in the region occupied by RdRp motifs A to E. There are also six positions of identity in region Y, containing the conserved amino acid motif LxxxxxGTFxQ. This motif is not conserved in RdRps of other RNA viruses and may be specific to this group. Pairwise comparisons of amino acid sequences using the GAP program, using either the whole sequence or just the conserved motifs A to E, showed that the A. thaliana ORF251 RdRp-like protein is more closely related to that of O. novo-ulmi Ld dsRNA 3a than to those of C. parasitica NB631 or R. solani 1A1 (Table 1). Particularly striking is the high degree of sequence similarity between motifs C of the O. novo-ulmi and A. thaliana RdRp-like proteins (Fig. 3). However, the putative A. thaliana RdRp-like protein was only about one third of the size of that encoded by Ld dsRNA 3a, corresponding to the conserved central region, and it is not known if the A. thaliana RdRp-like protein is expressed or functional.

On the basis of sequence similarities between RdRp-like proteins encoded by the Arabidopsis mitochondrial genome and fungal mitochondrial dsRNAs, Marienfeld et al. (1997) suggested the possibility of horizontal viral nucleic acid sequence transfer between fungi and plants. An alternative explanation is evolution of the sequences from a common dsRNA ancestor. Mitochondrial dsRNA elements have been described in land plants, e.g., alfalfa (Fairbanks et al., 1988), maize (Finnegan and Brown, 1986), sugarbeet (Powling, 1981), and green algae (Ishihara et al., 1992), and could be widespread. Furthermore, RdRp-like sequences, related to the fungal RdRp-like sequences, are encoded by the mitochondrial DNA of Vicia faba (Marienfeld et al., 1997), suggesting integration of such sequences into the mitochondrial DNA prior to separation of Arabidopsis and Vicia in plant phylogeny. It is also noteworthy that plant mitochondria use the universal genetic code (Gray, 1992), as do some groups of ancestral fungi, such as the chytridiomycetes and zygomycetes, which diverged prior to the ascomycetes and basidiomycetes (Paquin and Lang, 1996; Paquin et al., 1997). Sequences of a much wider range of fungal mitochondrial dsRNA-encoded RdRp-like proteins, including those of protist ancestors of fungi, of plant mitochondrial dsRNA-encoded RdRp-like proteins, and of plant mitochondrial DNAs, will be needed to distinguish between these possibilities.
Fungal mitochondrial virus-like RNAs belong to an evolutionary lineage distinct from those of most other fungal dsRNA and ssRNA viruses.

Viruses and virus-like RNA replicons are of common occurrence in fungi (Buck, 1986; Nuss and Koltin, 1990; Ghabrial, 1994) and may be encapsidated or unencapsidated. Several different families containing fungal viruses have been distinguished, the Totiviridae, the Partitiviridae, the Hypoviridae, the Barnaviridae, and probably the Reoviridae, in addition to the mitochondrial dsRNAs (reviewed by Buck, 1998). A dendrogram, derived from an alignment of amino acid sequences of the O. novo-ulmi Ld dsRNA 3a RdRp-like protein with those of other fungal RNA viruses, together with those of related animal, bacterial, and plant RNA viruses, is shown in Fig. 5. The fungal mitochondrial RdRp-like proteins, together with the Arabidopsis mitochondrial RdRp-like protein, form a cluster, strongly supported by the bootstrap analysis, which is distinct from the RdRp-like proteins of all the other fungal viruses which have been sequenced. In further analyses, no close relationships were found between the Ld 3a RdRp-like protein and RdRp-like proteins of dsRNA viruses of the Birnaviridae, Cystoviridae, or Reoviridae families, or animal and plant positive-stranded RNA viruses of RdRp supergroups II and III, as defined by Koonin (1991).

Ancestral relationships between RdRp-like proteins and RdRps encoded by the C. parasitica mitochondrial dsRNA, the T (23S) and W (20S) RNA replicons of Saccharomyces cerevisiae, and positive-stranded RNA bacteriophages of the Leviviridae family, reported by Polashock and Hillman (1994), can now be extended to include the O. novo-ulmi, R. solani, and A. thaliana mitochondrial RdRp-like proteins. This larger grouping was separated from all the other RdRps in 76% of the bootstrapped trees. The 23S and 20S ssRNA replicons of S. cerevisiae are present in great excess over the dsRNA forms (T and W, respectively) and may be regarded as positive-stranded RNA viruses (Wickner, 1992). The O. novo-ulmi Ld dsRNA 3a also exists in a ssRNA form which appears to be in excess over the dsRNA form and may also be regarded as a positive-strand RNA virus. Koonin and Dolja (1993) suggested that the ssRNA phases may have arisen from infection of bacteria with S. cerevisiae RNA replicons. However, the reverse scenario is also possible, i.e., introduction of ssRNA phages into fungi or their progenitors, followed by loss of the capsid and attachment proteins which may not be essential for fungal viruses in view of their intracellular mode of transmission. Indeed, since mitochondria are believed to have arisen from bacterial endosymbionts (reviewed by Gray, 1992), the fungal mitochondrial dsRNAs could have evolved from ssRNA phages of these endosymbionts. The S. cerevisiae 205 and 235 RNA replicons appear to replicate in the cytoplasm (not in the mitochondria) (Esteban et al., 1994) and may have arisen from an independent ssRNA phage introduction or an early virus RNA escape from the mitochondria.

Distant relationships have been reported between RdRp-like proteins of fungal and protozoal dsRNA viruses of the Totiviridae family (Bruenn, 1993), between those of the Totiviridae family and two unclassified dsRNAs, Agaricus bisporus L1 dsRNA (Van der Lende et al., 1996) and cucurbit yellows associated virus (CYAV) (Coffin and Coutts, 1995), between those of fungal and plant dsRNA viruses of the Partitiviridae family (Oh and Hillman, 1995; Nogawa et al., 1996), between those of fungal hypoviruses and plant positive-stranded RNA viruses of the Potyviridae family (Koonin et al., 1991), and between those of the mushroom bacilliform virus (Barnaviridae) and plant positive-stranded RNA viruses of subgroup II of the Luteovirus genus (Revill et al., 1994). All these relationships were well supported by the bootstrap analysis of the phenogram (Fig. 5). Koonin and Dolja (1993) placed the RdRps of viruses of the Potyviridae, Luteovirus subgroup II, the Hypoviridae, and the Totiviridae in supergroup I. The phenogram (Fig. 5) indicates that this supergroup could be extended to include RdRp-like proteins of the Partitiviridae family, the A. bisporus L1 dsRNA and CYAV. Two distinct lineages of virus RdRps in fungi are therefore indicated, one consisting of the supergroup I and related RdRps and a second consisting of RdRps of the fungal mitochondrial RNA replicons, the yeast RNA 205 and 235 RNA replicons, and the Leviviridae family.

MATERIALS AND METHODS

Fungal isolates

O. novo-ulmi isolates Log1/3-8d2 (Ld) and W2bOl1 have been described previously (Brasier, 1986). Single-conidial isolates, sci59, sci60, and sci64, derived from Ld, were obtained as described previously (Rogers et al., 1986). All fungal isolates were maintained on malt extract agar (MEA, Oxoid) as described (Rogers et al., 1986).

RNA isolation

Total nucleic acids were isolated as described (Covey and Hull, 1981) from mycelia after 14 days of growth on cellophane membranes overlaying MEA at 22°C. Ss and dsRNAs were separated by precipitating ssRNA with 2 M LiCl (Baltimore, 1966) and dsRNA with 5 M LiCl (Diaz-Ruiz and Kaper, 1978) as described previously (Hong et al., 1998). dsRNAs of isolate Ld were separated by PAGE (Rogers et al., 1986) and eluted from the gel using an RNaid kit (Bio 101).

cDNA synthesis, cloning, and sequencing

A cDNA library was constructed using gel-purified dsRNA. DsRNA, mixed with random hexamer primers (GIBCO-BRL), was denatured at 94°C for 3 min. First-
strand cDNA was then synthesised using Superscript II reverse transcriptase (GIBCO-BRL). Second-strand cDNA synthesis was as described (Gubler and Hoffman, 1983) using a Superscript Choice System for cDNA Synthesis (GIBCO-BRL). Double-stranded cDNA was ligated into the EcoRV site of the vector LITMUS28 (New England
Northern blotting

DsRNA and ssRNA fractions, pretreated with RNAse-free DNase (Promega) with or without S1 nuclease (GIBCO-BRL) according to the manufacturer's instructions, were separated by electrophoresis in a 14% agarose gel, denatured with 50 mM NaOH/150 mM NaCl, neutralized with 1 M Tris-HCl (pH 7.5)/15 M NaCl, and blotted to Hybond N membrane (Amersham). Blots were hybridized with strand-specific riboprobes prepared by in vitro transcriptions using T7 RNA polymerase (NBL) and [α-32P]UTP (Amersham) (Sambrook et al., 1989). DsRNAs were also analyzed by 8.0% PAGE prior to carrying out blotting and probing.

ACKNOWLEDGMENTS

We thank the Gatsby Charitable Foundation for support of this work and B. Müller for assistance. T.E.C. is the recipient of a CASE award from the Biotechnology and Biological Sciences Research Council.

REFERENCES


