© 1992 Oxford University Press
Nucleic Acids Research, Vol. 20, No. 15 3933–3937

*Foret*1, a reverse transcriptase-like sequence in the filamentous fungus *Fusarium oxysporum*

Jacqueline Julien, Simone Poirier-Hamon and Yves Brygoo
Institut de Génétique et Microbiologie, Unité associée au CNRS 1354, Université Paris-Sud, F-91405 Orsay Cedex, France

Received May 20, 1992; Revised and Accepted July 13, 1992
EMBL accession no. X65452

ABSTRACT

We report here the isolation of *Foret*1, a repeated DNA sequence cloned from the fungal plant pathogen *Fusarium oxysporum*. This clone exhibits a high degree of sequence similarity with the retroviral *pol* genes. Sequences homologous to protease, reverse transcriptase, ribonuclease H are found in that order. The overall structure is homologous to the 'gypsy' class of LTR-retrotransposons. Its similarity to elements present in widely different organisms may result from its horizontal transmission in recent evolutionary time.

INTRODUCTION

Transposable genetic elements have been studied extensively in yeast (1), *Drosophila* (2) and have been found in numerous other organisms subjected to molecular scrutiny (3). Detection and analysis of transposable elements in filamentous fungi have begun recently; most of them were discovered either as a consequence of mutational transposition (*Neurospora crassa* : 4) or as members of repetitive sequences (*Magnaporthe grisea*: 5) or fortuitously (*Podospora anserina* : 6; *Cladosporium fulvum*: 7).

The considerable variation and diversity observed in the plant pathogen *Fusarium oxysporum* led us to postulate that the presence of transposable elements might be responsible for a wide variety of genetic effects including chromosomal rearrangements and genome evolution. In our laboratory, the two quoted strategies have been developed to trap transposable elements. Daboussi and coworkers have successfully used the inactivation of the *nia*D gene to isolate a new family of fungal transposable elements (8). The analysis of dispersed repetitive DNA sequences has allowed us to clone a reverse transcriptase-like element. We report here the isolation of p2481, a DNA sequence which exhibits homology to viral and retrotransposon sequences and therefore represents a newly discovered member of this class of elements.

MATERIALS AND METHODS

Fungal strains and culture media

The strain of *Fusarium oxysporum f.sp. melonis* used is Fom 24 belonging to race 1,2 (9) and was obtained from the Station de Recherches sur la flore pathogène dans le sol, INRA, Dijon, France. It was maintained on PDA (potato-dextrose agar) (10).

For DNA isolation, Fom 24 was grown 48 h at 26°C in Roux flasks containing 100 ml of potato-dextrose broth supplemented with 0.1% (w/v) yeast extract and 0.1% casein hydrolysate (Difco, Detroit, USA).

Bacterial strains

Escherichia coli DH5α was used for construction, propagation and amplification of recombinant plasmids as described by Sambrook and coworkers (11).

Isolation and manipulation of DNA

Fom 24 DNA was isolated as described by Daboussi and coworkers (12). Plasmid DNA was purified by the method of Holmes and Quigley (13). Restriction enzyme digestions, ligations, electrophoreses, *in vitro* ^{32}P-labelling of plasmid DNA by nick-translation (Amersham kit) and membrane hybridizations with ^{32}P labelled DNA probes were performed following standard procedures (11). DNA transfer to nylon membranes (Hybond N, Amersham UK) was achieved using a vacuum blotting system (Pharmacia, Uppsala, Sweden).

Construction of minilibraries

Fom 24 total DNA was digested with *Xho* I and electrophoresed on a low-melting point agarose gel. Under U V light, six adjacent slices of agarose were cut out from the gel covering the range from 8 to 3 kb . Each slice, containing DNA fragments differing by approximatively 1 kb, was separately electroeluted and inserted by ligation into *Sal* I-cut pUC18.

Minilibraries were obtained after transformation of *E.coli* DH5α by these sets of ligation products.

DNA Sequencing

DNA sequences were determined by the dideoxyribonucleotide chain termination method (14) using modified T7 polymerase (Pharmacia kit) (15). Either universal primers or primers deduced from sequencing and synthesized on an Applied Biosystems Apparatus 381 A were used.

The Fom 24 sequence was compared with those of various retroid elements. For each domain, the amino acid sequences were aligned by eye and each one was compared to the others using the MULTALIN program (16). The nucleotide sequence reported in this paper appears in the EMBL, Genbank and DDBJ Nucleotide sequence Databases under the accession number X65452.

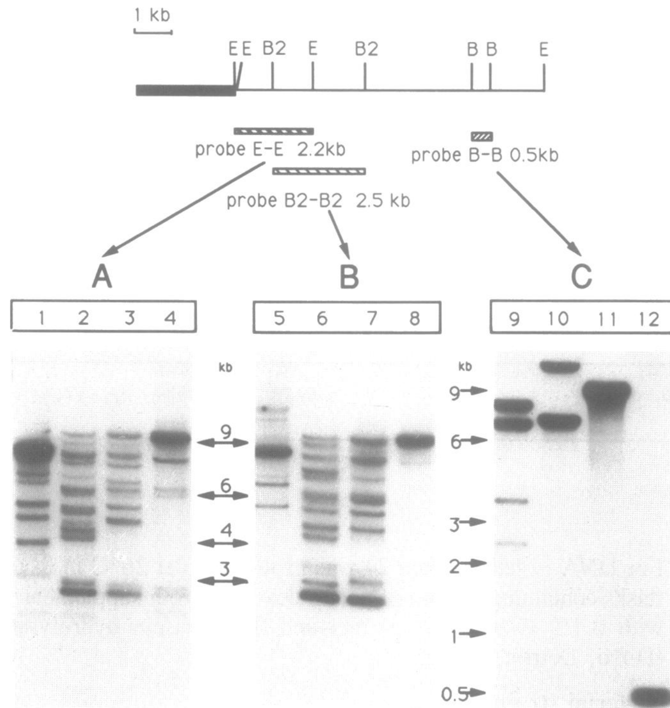

Figure 1. Upper part: Position of relevant restriction sites of the plasmid p2481. The symbols used to represent the restriction enzymes are: E=*Eco* RI, B=*Bam* HI and B2= *Bgl* II. Boxed area represents pUC18 sequence. Lower part: Southern blot of genomic DNA of Fom 24 probed with the fragments represented by the hatched boxes and obtained after enzymatic digestion of the plasmid DNA, electroelution and labelling by nick-translation. Genomic DNA was digested by *Cla* I (lanes1,5,9), *Cla* I and *Bgl* II (lanes 2,6,10), *Bgl* II (lanes 3,7,11), *Xho* I (lanes 4,8), *Bam* HI (lane 12).

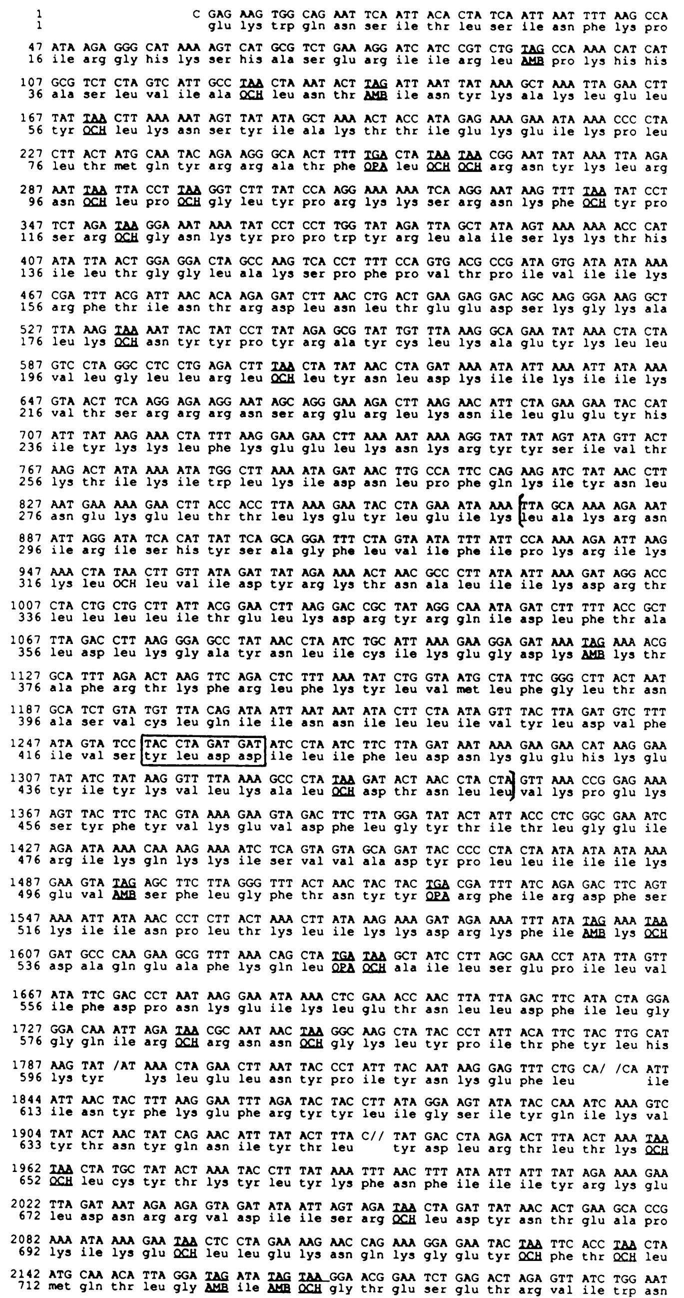

Figure 2. The DNA sequence of the cloned *Eco* RI-*Eco* RI 2.2kb. The amino acid sequence is shown below the proposed reading frame. Stop codons are underlined. The RT domain is bracketed and the tyr-X-asp-asp motif boxed. Frameshifts indicated by slashes (/ at nucleotide1793,1839 and 1936) are proposed to gain a better alignment in Fig.4.

RESULTS

Cloning of dispersed repetitive DNA sequences

When Fom 24 total DNA digested with different restriction endonucleases was run on a 0.8% agarose gel and stained with diluted ethidium bromide solutions, a few discrete bands became visible over a smear indicating the presence of repeated fragments. In filamentous fungi there are few classes of repeated DNA sequences, the ribosomal DNA multigene family being the most represented. Since our aim was to clone non ribosomal repeated sequences, we first located rDNA among the fragments. Southern blots of Fom 24 DNA restriction digests were probed with either ^{32}P- labelled Fom 24 ribosomal RNA or ^{32}P nick-translated plasmid pPN2 containing the ribosomal unit of *Podospora anserina* (17). This experiment located the ribosomal restriction fragments for each restriction endonuclease (data not shown). The digestion of DNA by *Xho* I appeared to be the most appropriate because the ribosomal probes showed hybridization to a single band located at a size about 9–10 kb whereas other repeated bands corresponding to sizes of 8, 5.5, 4.5 and 3 kb respectively were detected under UV. In order to clone the DNA from these 4 bands, minilibraries were constructed as described in Materials and Methods.

The minilibrary containing the 8 kb fragments provided several recombinant clones. Each recombinant plasmid was extracted and tested for the presence of repeated DNA sequences by probing a Southern blot of Fom 24 genomic DNA. One of them, p2481, which yielded an obvious repetitive banding pattern, was selected for further study. In order to analyse this more precisely, genomic DNA was probed with three internal fragments of the insert as depicted in Figure1. In parts A and B, the probes used were two fragments isolated from the left side of the insert. For various enzymatic digestions, both probes displayed a complex banding pattern consistent with related sequences being present at a variety of sites within the genome. Different band intensities may be due to sequence variation between some related sequences; nevertheless, in each lane the strongest signal is located at a size

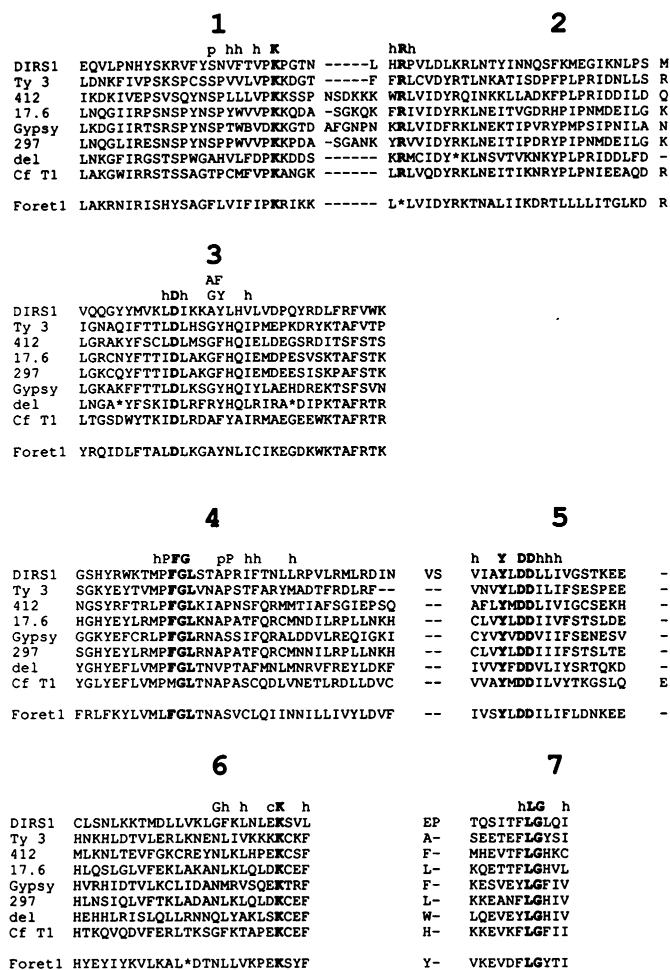

Figure 3. Optimal alignment of inferred *Foret*1 RT amino acid sequence with other retroid elements shown in N to C terminus order. The single letter code for amino acids is used. Other sequences are from: DIRS-1 (20) from *Dictyostelium discoideum*; Ty 3 (21) from *Saccharomyces cerevisiae*; 412 (22), 17.6 (23), Gypsy (24), 297 (25) from *Drosophila*; del (26) from *Lilium henryi*; *Cf* T1 (7) from *Cladosporium fulvum*,. Numbers 1 to 7 refer to domains 1−7 identified by Xiong and Eickbush. Largely unvaried or chemically similar residues are shown at the top of the alignment: h, hydrophobic residue; p, small polar residue; c, charged residue. Amino acids in bold type are invariant among presented RTs . An asterisk (*) indicates a stop codon.

expected from the restriction map. On the contrary, the probe used in part C revealed only one or two bands (depending on the restriction enzyme used) indicating an absence of polymorphism for this part of the insert .

DNA sequencing and analysis of deduced polypeptide products

These results led us to carry out the sequencing of the 2.2 kb *Eco* R1- *Eco* R1 fragment located on the left end of the insert in p2481. The nucleotide sequence obtained (with the corresponding amino acid sequence in the best reading frame) is given in Figure 2: striking features are the presence of numerous stop codons and the identification of the amino acid motif tyr-X-asp-asp (nucleotides 1256 to 1267) known to be characteristic for polymerases (18). This led us to undertake a comparison to retroid elements.

Table 1. Identity percentage of amino acid sequences in *Foret*1 regions with those in other retroid elements.

Element	RT%	RNase H%
Ty 3	35	30
412	27	27
17.6	33	36
Gypsy	34	26
297	35	28
del	36	30
Cf T1	44	38
DIRS-1	25	20

The covered domains are for RT, 178 amino acids (domains 1−7 fig.3); for RNase H, 109 amino acids.

Reverse transcriptase

The hypothetical polypeptide deduced from the sequence from nucleotide 872 to nucleotide 1411 ignoring two stop codons (nucleotides 1118 and 1334) was found similar to retroviral reverse transcriptases (RT). Xiong and Eickbush (19) have made extensive comparisons of RT of various origins and identified seven blocks of amino acid homology that appear to be diagnostic for RT. In the sequence which we isolated from Fom 24 and called *Foret*1(for *Fusarium oxysporum* retroid element), all seven of these blocks of homology are present with the appropriate spacing and with the appropriate location (Figure 3). Alignment of the amino acid sequence of *Foret*1 and those of eight representative elements of the LTR-containing retrotransposon branch is also presented in Figure 3. Close conservation in order and spacing of the seven motifs is observed. In addition, the 178 amino acid sequence of *Foret*1 was compared pairwise with those of members of the gypsy group (Table 1: DIRS-1 sequence does not fit in the gypsy group and is used as an outer reference). The percentage of identity ranges from 27% to 44% which is significant for related sequences (27). This lends support to the proposal that *Foret*1 is a member of the LTR-retrotransposon branch and more precisely of the gypsy group. It is noticeable that the most related sequences *Cf* T1 (44%), del (36%), 297 (35%) originate from fungi,plants and animals, respectively.

Ribonuclease H

Doolittle and coworkers (3) have published evolutionary trees based on the comparison of the amino acid sequences of retroviral proteins. RNase H sequences are estimated to change at a faster rate than RT ones. The phylogeny determined by these authors from the alignment of 26 RNase H sequences is consistent with the gypsy group previously defined and gives prominence to remarkable resemblances in amino acid sequences between the RNase H from *E.coli* and the RNase H domain of many retroelements. Figure 4 presents the best alignment of the RNase H amino acid domain of *Foret*1 (from nucleotide 1412 to nucleotide 2087 in fig. 2) with those from elements of the gypsy group. Amino acid identities between *Foret*1 and the gypsy family are easily observed in the regions close to C and N termini of the sequence (first and last rows). The percentage of identity calculated for the 120 first amino acids downsteam of the RT (first and second rows in fig.4) is given in Table 1. For five out six stop codons in this region, a single nucleotide change would substitute an amino acid present in the corresponding *Cf* T1 sequence. In order to gain a better alignment of the putative

```
Ty 3     GIQKIAPLQHKCAAIRDFPTPKTVKQAQRFLGMINYYRRFIPNCSKIAQPIQLFI-----
412      TDKGILPDDKKYDVIQNYPVPHDADSARRFVAFCNYYRRFIKNFADYSRHITRLCK----
17.6     TPDGIKPNPEKIEAIQKYPIPTKPKEIKAFLGLTGYYRKFIPNFADIAKPMTKCLK----
Gypsy    SKDGTKSDPEKVKAIQEYPEPDCVYKVRSFLGLASYYRVFIKDFAAIARPITDILKGENG
297      TPDGIKPNPIKVKAIVSYPIPTKDKEIRAFLGLTGYYRKFIPNYADIAKPMTSCLK----
del      SREGIVVDPVKVKAVMNWELPKNIFEIRSFLGLAGYYRRFIKGFSKLAALMT--------
CfT1     STTGITIDPAKTQSIREWPEPKTVKDVQSFLGLANYNRKFIKDYSKTAAPMT--------

Foret1   TLGEIRIKQKKISVVADYPLLIIIKEVqSFLGFTNYYrRFIRDFSKIINPLTKLIK----


Ty 3     --------CDKSQWTEKQDKAIDKLKDALCN-SPVLVPFNNKANYRLTTDASKDGIGAVL
412      -------KNVPFEWTDECQKAFIHLKSQLINPT-LLQYPDFSKEFCITTDASKQACGAVL
17.6     --KNMKIDTTNPEYDS----AFKKLKY-LISEDPILKVPDFTKKFTLTTDASDVALGAVL
Gypsy    SVSKHMSKKIPVEFNETQRNAFQRLRNILASEDVILKYPVFKKPFDLTTDASASGIGAVL
297      --KRTKIDTQKLEYIE----AF--LKA-LIIRDPILQLPDFEKKFVLTTDASNLALGAVL
del      ---QLTKKGENFNWTKKYQNSFDELKRXLTTV-PVLTIPISG*PFVVYTDASLAGLEGVL
CfT1     ---MLTAKDVNWKWGKEQTEAFKRLKEQCASA-PTLRLFDGSKEVHIETDASDMAIGACL

Foret1   -------KDRKFIwKkDAQEAFKQL*eAILSEPILVIFALIRKeNSKPTY*TSY*EDK-L


Ty 3     EEVDNKNKLVGVVGYFSKSLSSAQKNYPAGELELLGIIKALHHFRYMLHGKHFTLRTDHH
412      TQNHNGHQLPVAYASRAFTKGESNKSTTEQELAAIHWAIIHFRPYIYG-KHFTVKTDHRP
17.6     SQ--DGH--PLSYISRTLNEHEINYSTIEKELLAIVWATKTFRHYLLG-RHFEISSDHQP
Gypsy    SQ--EGR--PITMISRTLKQPEQNYATNERELLAIVWALGKLQNFLYGSREINIFTDHQP
297      SQ--NGH--PISFISRTLNDHELNYSAIEKELLAIVWATKTFRHYLLG-RQFLIASDHQP
del      MQ--DGR--VVAYASRQLKVHENNYPTHDLELAVVIFILKLWRHYLGG-EDFELYCDHKS
CfT1     TQTHDGKRHPVAYYSAKMTTAEQNYDIHDKELLAIVAAMQHWRVYVEGPPKLTILSDHKN

Foret1   DNAITKASYTLLHSTCISI*TRTyLPYLQ*GVSAGLLTTLRNLDTTL*EVYTKSKSILTI


Ty 3     ISLLSLQNKNEPARRVQRWLDDLATYDFTLEYLAGPKNVVADAISRAVYTITPETSR
412      LTY--LFSMVNPSSKLTRIRLELEEYNFTVEYLKGKDNHVADALSRITIKELKDIT-
17.6     LSW--LYRMKDPNSKLTRWRVKLSEFDFDIKYIKGKENCVADALSRIKLEETYLSEQ
Gypsy    LTF--AVADRNTNAKIKRWKSYIDQHNAKVFYKPGKENFVADALSRIQNLNALQNP-
297      LRW--LHNLKEPGAKLERWRVRLSEYQFKIDYIKGKENSVADALSRIKIEENHHSE-
del      LKY--ISTQKDLNLRQR*WIEVLKDFDFSIFYHPGKANVVADALSRKSQISHLISAR
CfT1     LTY--FTTTKELTRRQARTSELLGQYKFEIKYTPGTENGPADALSRRSDYMEGKEPV

Foret1   RTFILYYDLRTLTK*qLCYTKYLYKFNFIIIYRKELDNRRVDIISRrLDYNTEAPKI
```

Figure 4. Optimal alignment of sequences corresponding to RNase H domain shown in N to C terminus order from retroid elements presented in Fig.3 (except DIRS-1 which does not fit in the gypsy group). Closed circles indicate residues which are crucial for the activity of *E.coli* RNase H and which are well conserved in most RNase H. Closed triangles are conserved hydrophobic and aromatic groups. Amino acids in bold type are invariant among sequenced RNase H . An asterisk (*) indicates a stop codon. A slash (/) indicates a frameshift. Lower-case letters are used when a single mutation allows to replace a stop codon by the amino acid present in the *Cf* T1 sequence.

```
17.6     ENNLKCLIDTGSTVNMTSKNIFD-LPIQ---NTSTF----IHTSNGPLIVNKSIIIPSKI
del      SSLCHVLIDTGSTHSFITPRIIKMLEIPV--QPLGYILSVISPIGTSTFVNQVCKGCMIT
Cf T1    GHKTDAMIDSGASGNFASESFVTRNRIATCKKKEGYELIAVDGSSLPSVERETIPLPL-A

Foret1   MQYRRATF*L**RNYKLRN*LP*GL-YPRKKSRNKF*YPSR*GNKYPPWYRLAI-----S


17.6     LFPTTNEFLLHPFS-ENYDLLLGRKLLAEAKATISYRDQEVTLY-------NNKYKLIEG
del      IGNQELTVDLIILDLEDPDILLGMDWLAAYHVVLDCFSKKVTFH----LPGIPEFHFH--
Cf T1    IQRHHEEITLDVTDMASHDIVLGMPTLRKHNPVIDTRRGVLTFRECECVIDIQPAQTQRS

Foret1   KKTHILTGGLAKSPFPVTPIVI-IKRFTINTRDLNLTEEDSKGK----ALK*NYYPYR-A


17.6     IATHEQSHFQNVNMIPDTMLRQPNKISPILESDLYRLEHLN--------NEEKQRLCALL
del      --GETQHTFFRT-FTHQPNLSYLASLASEINITPSTDLSL------------------IV
Cf T1    LADEARKQLNRIQLAPTRRTEEPPSTGTDTGVGPPGHGVTGSDGSNAPSKDTNISELSIP

Foret1   YCLKAEYKLLVLGLLRL-*LYNLDKIIKIIKVTSRRRNSRERLK------------NIL


17.6     QKYHD--IQYHEG---DKLTFTNQTKHTINTKH-NLP-LYSKYSYPQAYEQEVESQIQDM
del      REYIN---VFDDDLP--GLPPPREIEFQINLLPGTSPISITPYHMAPSELQELKEQLEDL
Cf T1    KEYRKTSRLFEEERGKDALPKHQPTDHKINIQPGKEPPTGPLYQMSEKELQTLRETLKEK

Foret1   EEYHIYKKLFKEELKNKRYYSIVTKTIKIWLKIDNLP-FQKIYNLNEKELTTLKETLEIK
```

Figure 5. Alignment of sequences corresponding to protease domain shown in N to C terminus order from *Foret*1 and 3 elements presenting the best similarity for the RT domain. Amino acids in bold type are invariant in the four presented proteases. Lower-case letters are used when a single mutation allows to replace a stop codon by the amino acid present in the *Cf* T1 sequence.

product, it is necessary to introduce three frameshifts (see legend fig2). Numerous stop codons as well as frameshifts may be the indication of *in vivo* mutations in a formerly active element. Structural models of ribonuclease H from retroviruses were built by Nakamura and coworkers (28). These authors concluded that the essential amino acids for the activity in *E.coli* RNase H are not only well conserved in all sequences but residues constructing a hydrophobic core are also located at the same position as those of *E.coli* RNase H (see fig.4). Therefore *Foret*1 contains sequences related to RNase H particularly in the regions close to C and N termini of the sequence even though key motifs have apparently disappeared.

Protease

The protease polypeptide product corresponds to the translation from the DNA stretch between nucleotide 233 and nucleotide 871 (figure 2). The *Foret*1 sequence exhibits nine stop codons. For two out of the nine stop codons, a single nucleotide change would substitute an amino acid present in the corresponding *Cf* T1 sequence. An attempt to align that amino acid sequence with those of the gypsy group is presented in figure 5. There is no easily recognized protease-like sequence. This is not surprising as the protease is one of the fastest changing proteins in retroviruses (3). Sequence homology can be detected only in the C terminus region directly upstream of the RT region. The best fit is between *Foret*1 and *Cf* T1 (20 out of 60 amino acids are identical).

DISCUSSION

Fusarium oxysporum infects over 100 botanical species. The cloned *Foret*1 element is probably a significant part of the panoply of mobile elements that give rise to repeated sequences and cause important evolutionary variation including the ability of *Fusarium oxysporum* to overcome novel sources of plant resistance. Experiments are in progress in which DNA isolated from various strains is probed with the *Foret*1 element. That will give us data for its occurrence in different *formae speciale* and information on a possible role in variability.

The order in which catalytic functions are encoded in LTR-retroelements divides the group into two subsets (19). The first one corresponds to the arrangement protease, integrase, reverse transcriptase, ribonuclease H found in copia of *Drosophila*, Ta 1 of *Arabidopsis thaliana*, Tnt1 of *Nicotiana tabacum*, Ty1, Ty2 of *Saccharomyces cerevisiae.* The second one corresponds to the arrangement protease, reverse transcriptase, ribonuclease H, integrase found in the gypsy group of *Drosophila*, Ty3 of *Saccharomyces cerevisiae*, del of *Lilium henrii* and *Cf* T1 of *Cladosporium fulvum. Foret*1 and *Cf* T1 are the filamentous fungi elements in this set. Integrase was not identified in the sequenced 2.2kb *Eco* R1-*Eco* R1 fragment of *Foret*1; its location on the plasmid p2481 can be predicted adjacent to the right *Eco* R1 site and will be sequenced in the future.

Accumulation of stop codons in required coding regions observed in *Foret*1 has been already described for some clones corresponding to individual retroviral-like elements belonging to families where the copy numbers are quite variable such as del from *Lilium henryi* (26), Tad from *Neurospora crassa* (29) and SURL from sea urchin (30). This suggests that the inferred elements are remnants of ancient functions or pseudogene sequences. As reviewed recently in higher plants (31) only a very small number of plant retroelements have been shown to be active. In *Fusarium oxysporum*, the study of other individual clones will give information on the dispersed elements present in the same genome and lead to a possible isolation of an active

element if it is present. In *Cladosporium fulvum*, full-length copies of *Cf* T1 element and virus-like particles (VLP) have been observed but there is no direct evidence that the elements are capable of further transposition (32).

Emergence of LTR-elements from divergent host species: yeast, insects, plants indicates that either the progenitors of these groups of elements are very ancient or there has been extensive horizontal transfer of elements between species implying lost functions or independant vectors. Also indicative of horizontal transfer is the grouping reported for *Arabidopsis thaliana* where LTR-elements closely related to insect ones coexist with other distinct LTR-elements in the same plant species, suggesting a recent extrachromosomal origin for the first set of elements (31). In case of the retrotransposons, the presence of VLP would provide favourable vectors for possible horizontal transfers involving plants and others organisms such as insects or fungi that are evolutionary distant but ecologically related.

ACKNOWLEDGEMENTS

We would like to thank J.F.Lafay for help with computer analysis and C.Nauche for some experiments. We are grateful to R.P.Oliver for making unpublished results available.

REFERENCES

1. Boeke, J. D.(1989) in Mobile DNA(Berg D.E. and Howe M. Eds) American Society for Microbiology Washington pp335–374.
2. Bingham P.M. and Zachar, Z. (1989) in Mobile DNA Berg D.E. and Howe M. (Eds) American Society for Microbiology Washington pp485–502.
3. Doolittle, R.F., Feng, D.F., Johnson, M.S. and McClure, M.A. (1989). Quant. Rev. Biol. **64** , 1–30.
4. Kinsey, J.A. and Helbert, J. (1989) Proc. Ntl. Acad.Sci. USA **86**, 1929–1933.
5. Valent, B. and Chumley, F. G. (1991) Annu. Rev. Phytopath., **29**, 443–447.
6. Deleu, C., Turcq, B. and Begueret, J. (1990) Nucl. Acids Res. **18** 4901–4903.
7. McHale M.T., Roberts, I.N., Talbot, N.J. and Oliver, R.P. (1989). Molecular-Plant-Microbe Interactions **2**, 165–168.
8. Daboussi, M. J., Langin, T. and Brygoo, Y. (1992) Mol. Gen. Genet. **232**, 12–16.
9. Bouhot, D. (1981). . In Nelson, P.E., Toussoun, T.A., Cook, R.J. (Eds), *Fusarium*, Diseases, Biology and Taxonomy. The Pennsylvania State University Press University Park PA, pp. 318–326.
10. Daboussi-Bareyre, M.J. (1980). J. Gen. Microbiol. **116** , 425–433.
11. Sambrook J., Fritsch, E. and Maniatis, T. (1989). Molecular cloning: A Laboratory Manual. Cold Spring Harbour Press , New York: Cold Spring Harbor Laboratory.
12. Daboussi, M.J., Djeballi, A., Gerlinger, C., Blaiseau, P.L., Bouvier, I., Cassan, M., Lebrun, M.H., Parisot, D., Brygoo, Y. (1989). Curr. Genet. **15** , 453–456.
13. Holmes, D.S. and Quigley, M. (1981). Anal. Biochem. **114** , 193.
14. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad.Sci. USA **74** , 5463–5467.
15. Tabor, S. and Richardson, C.C. (1989) J. Biol. Chem. **264**, 6447–6458.
16. Corpet, F. (1988) Nucl. Acids Res. **16**, 10881–10890.
17. Labat, N., Perrot, M. and Begueret, J. (1985). Mol. Gen. Genet. **199** , 154–157.
18. Toh, H., Hayashida, H., Miyata,T. (1983) Nature **305**, 827–829.
19. Xiong, Y. and Eickbush, T.H.(1990) EMBO J. **9** , 3353–3362.
20. Cappello, J., Handelman, K. and Lodish, H.F. (1985). Cell **43**, 105–115.
21. Hansen, L.J, Chalker, D.L. and Sandmeyer, L.B. (1988) Mol. Cell. Biol. **8**, 5245–5256.
22. Yuki, S., Inouye, S., Ishimaru, S. and Saigo, K.(1986) Eur. J. Biochem. **158**, 403–410.
23. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. and Yuki, S. (1984) Nature **312** , 659–661.
24. Marlor, R.L., Parkhurst, S.M., Corces, V.G. (1986). Mol. and Cell. Biology, **6** 1129–1134.
25. Inouye, S., Yuki, S. and Saigo, K. (1986). Eur. J. Biochem., **154** , 417–425.
26. Smyth, D.R., Kalistis, P., Joseph, J.L. and Sentry, J.W. (1989). Proc. Natl. Acad. Sci. USA **86** , 5015–5019.
27. Doolittle, R.F.(1986). Of URFs and ORFs: A Primer on How to analyse derived amino acid sequences. Oxford University Press.
28. Nakamura, H., Katayanagi, K., Morikawa, K., and Ikehera M, (1991) Nucl. Acids Res.**19**, 1817–1823.
29. Kinsey, J. A. (1990) Genetics **126**, 317–323.
30. Springer M.S., Davidson, E. H. and Britten, R. Y. (1991) Proc. Natl. Acad. Sci. USA **88** , 8401–8404.
31. Grandbastien, M. A. (1992) Trends in Genetics **8**, 106–108.
32. McHale, M.P., Roberts, I.N., Noble, S.M., Beaumont, C., Whitehead, M.P., Devanshi, S., Oliver, R.P. (1992) submitted.