

Pyret, a Ty3/Gypsy retrotransposon in *Magnaporthe grisea* contains an extra domain between the nucleocapsid and protease domains

Hitoshi Nakayashiki*, Hideki Matsuo, Izumi Chuma, Kenichi Ikeda, Shigeyuki Betsuyaku, Motoaki Kusaba¹, Yukio Tosa and Shigeyuki Mayama

Laboratory of Plant Pathology, Faculty of Agriculture, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan and ¹Laboratory of Plant Pathology, Faculty of Agriculture, Saga University, 1 Honjo-cho, Saga 840-8502, Japan

Received July 23, 2001; Revised and Accepted September 3, 2001

DDBJ/EMBL/GenBank accession no. AB062507

ABSTRACT

A novel Ty3/Gypsy retrotransposon, named Pyret, was identified in the plant pathogenic fungus *Magnaporthe grisea* (anamorph *Pyricularia oryzae*). Pyret-related elements were distributed in a wide range of *Pyricularia* isolates from various graminaceous plants. The Pyret element is 7250 bp in length with a 475 bp LTR and one conceptual ORF. The ORF contains seven nonsense mutations in the reading frame, indicating that the Pyret clone is lightly degenerate. Comparative domain analysis among retroelements revealed that Pyret exhibits an extra domain (WCCH domain) beyond the basic components of LTR retrotransposons. The WCCH domain consists of ~300 amino acids and is located downstream of the nucleocapsid domain. The WCCH domain is so named because it contains two repeats of a characteristic amino acid sequence, W-X₂-C-X₄-C-X₂-H-X₃-K. A WCCH motif-like sequence is found in the precoat protein of some geminiviruses, viral RNA-dependent RNA polymerase and also in an *Arabidopsis* protein of unknown function. Interestingly, detailed sequence analysis of the gag protein revealed that Pyret, as well as some other chromodomain-containing LTR retrotransposons, displays significant sequence homology with members of the gammaretroviruses (MLV-related retroviruses) in the capsid and nucleocapsid domains. This suggests that chromodomain-containing LTR retrotransposons and gammaretroviruses may share a common ancestor with the gag protein.

INTRODUCTION

Long terminal repeat (LTR) retrotransposons show an overall organization strikingly similar to retroviruses, which consists

of gag, pol and, in some cases, env genes flanked by LTRs at each end. In the most recent virus taxonomy it was proposed that LTR retrotransposons are classified in the Retroelementopsida, the class to which retroviruses belong, and are further divided into two main families, Pseudoviridae (corresponding to the Ty1/Copia subgroup) and Metaviridae (Ty3/Gypsy subgroup) (1,2). The phylogenetic relationships and origin of these retroelements are of great interest because they are ubiquitous components of eukaryotic genomes and thereby are supposed to have played important roles in the evolution of the genome (3,4). Most phylogenetic studies of retroelements have been conducted based on the pol gene, which is the most highly conserved among the retroelement genes, revealing that LTR retrotransposons have a discrete phylogenetic lineage from retroviruses (3,4).

On the other hand, sequence homology of gag proteins is not generally found among distantly related retroelements (5). Therefore, phylogenetic analysis of gag proteins is rarely reported. The gag genes of simple retroviruses typically encode the matrix (MA), capsid (CA) and nucleocapsid (NC) proteins (5). The gag proteins of some retroviruses, however, encode one or more additional domains that vary among different viruses. Retroviral MA protein is required for membrane targeting of the gag protein and for capsid assembly. The CA protein forms the prominent hydrophobic core of the viral virion (6) and the NC domain encodes a small basic protein that plays a critical role in the RNA selection leading to RNA packaging into the viral particle (7,8). The NC protein contains one or two zinc finger motifs characterized by the amino acid sequence C-X₂-C-X₄-H-X₄-C, which is considered to be a hallmark of the gag protein in retroviruses and also LTR retrotransposons, with some exceptions. In addition, a stretch of ~20 amino acids in the CA domain is highly conserved in most retroviruses, except for spumaretroviruses, and is termed the major homology region (MHR) (9). An MHR-like motif has also been identified in an LTR retrotransposon, Ty3, in its CA domain and was shown to be involved in multiple steps of Ty3 retrotransposition (10).

Despite the information obtained in retroviruses, very little is known about the gag proteins of most LTR retrotransposons

*To whom correspondence should be addressed. Tel: +81 78 803 5867; Fax: +81 78 803 5867; Email: hnakaya@kobe-u.ac.jp

because only limited sequence homology has so far been described between retroviruses and LTR retrotransposons. In this paper, we isolate a new LTR retrotransposon named Pyret from the phytopathogenic fungus *Magnaporthe grisea* and perform a detailed sequence analysis of Pyret, especially the gag protein. Comparative sequence analysis revealed that this element is novel because it carries an extra domain between the NC domain and protease, which is characterized by a short stretch of specific amino acid sequence, termed the WCCH motif.

MATERIALS AND METHODS

Fungal strains and growth conditions

The original hosts and locality of fungal strains used to investigate the distribution pattern of Pyret-related elements can be found in Eto *et al.* (11) except for a *Pyricularia zingiberi* isolate from *Zingiber mioga*, HYZiM101-1-1-1. The isolates were purified by monoconidial isolation and maintained on PDA medium for short-term storage or on sterilized barley seeds for long-term storage, as described previously (12). For DNA extraction, fungal mycelia were grown in CM liquid broth (0.3% casamino acids, 0.3% yeast extract, 0.5% sucrose) at 26°C for 5 days on an orbital shaker (120 r.p.m.).

Construction of a genomic library

The cosmid vector pMLF2 was kindly provided by Dr Sally A. Leong. A genomic library of GFSI1-7-2, a foxtail millet isolate of *M. grisea*, was constructed in pMLF2. Fifty micrograms of total DNA was partially digested with *Sau3AI* and ligated into the *Bam*HI-digested pMLF2 vector.

DNA isolation and analysis

Total fungal DNA was isolated as described previously (12). DNA gel blot analysis was performed using a dioxetane chemiluminescence system, Gene Images (Amersham, Arlington Heights, IL). *Eco*RI-digested fungal DNA was separated on a 0.8% TAE-agarose gel and transferred to a nylon membrane. A 0.6 kb DNA segment used for probing Pyret-related elements, corresponding to the gag domain (nt 1307–1902) of Pyret, was cloned from a *M. grisea* isolate (NNPM3-1-1) and then labeled with fluorescein by the random prime labeling method. Hybridization was performed in 5× SSC, 0.1% (w/v) SDS, 5% (w/v) dextran sulfate, 5% (v/v) liquid block (Amersham) at 60°C overnight. Southern blots were then washed twice in 1× SSC containing 0.1% SDS for 15 min at 65°C and twice in 0.5× SSC containing 0.1% SDS for 15 min at 65°C. Detection procedures were performed according to the manufacturer's instructions.

DNA sequencing and analysis

Sequencing reactions were performed using the ABI Prism Big-dye Terminator Ready Reaction sequencing kit (Applied Biosystems, Foster City, CA) and analyzed with an ABI310 sequencer. Sequence searches were conducted using the BLAST program (13) available at DDBJ (<http://www.ddbj.nig.ac.jp/>). Amino acid sequences were aligned with the CLUSTAL W program (14) using the BLOSUM series of protein weight matrices (15) and adjusted manually if necessary using Jalview (<http://www.ebi.ac.uk/~michele/>).

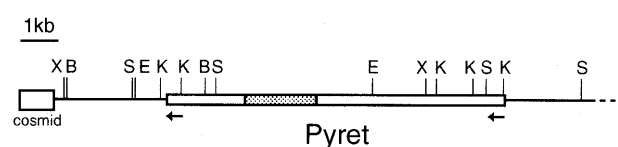


Figure 1. Partial structure of the cosmid clone S8-C-6. Arrows represent long terminal repeats. The shaded box indicates an MGR586-like element inserted in Pyret. The nucleotide sequence of Pyret has been deposited in the DDBJ/EMBL/Genbank nucleotide databases under accession no. AB062507. B, *Bam*HI; E, *Eco*RV; K, *Kpn*I; S, *Sac*I; X, *Xho*I.

jalview/). A phylogenetic tree was constructed by the neighbor-joining method of Saito and Nei (16) and viewed using TreeView (17). The tree was subjected to bootstrap resampling of 1000 replicates.

RESULTS

Isolation of Pyret from a foxtail millet isolate of *M. grisea*

A clustered distribution of transposons has been often reported in *M. grisea* (18–20). To analyze a transposon cluster in the *M. grisea* genome, we screened a cosmid library of GFSI1-7-2, a foxtail millet isolate of *M. grisea*, by probing with five known *M. grisea* transposable elements including MAGGY, Pot2, MGR583, MGR586 and Mg-SINE. Cosmid clone S8-C-6, which hybridized to four of the five transposons, was chosen for further sequence analysis. The insert of S8-C-6 was ~40 kb long. A BLAST database search allowed us to detect a 9 kb long retrotransposon-like sequence delimited by two 475 bp LTRs in a partial sequence of the S8-C-6 insert (Fig. 1). The retrotransposon-like sequence contained a 1.8 kb insertion of an MGR586-like DNA transposon. The conceptual translation product of the retrotransposon-like sequence exhibits a significant degree of similarity to gene products of several Ty3/Gypsy LTR retrotransposons. Therefore, we concluded that this sequence corresponds to a new Ty3/Gypsy LTR retrotransposon of *M. grisea*, which we named Pyret.

Sequence analysis revealed that Pyret is 7250 bp long and contains one putative open reading frame (ORF) composed of 1811 amino acid codons including seven nonsense mutations. The 5'- and 3'-LTR sequences of Pyret are 93.5% identical with 31 base substitutions. Interestingly, all the base substitutions are C:G→T:A transitions that are characteristic of repeat-induced point mutation (RIP), which was originally found in *Neurospora crassa* (21). Hence, the Pyret copy isolated in this study is lightly degenerate, probably by a RIP-like process as reported in another *M. grisea* retrotransposon, MAGGY (22). The Pyret LTR sequence begins and ends with the 7 bp terminal inverted repeat sequences TGTTACG...CGTAACA that are highly conserved in LTRs of several fungal retrotransposons such as Cft-1, Grasshopper, Skippy, MGLR-3, Boty and Cgret (23–28). Pyret is flanked by the 5 bp direct repeat sequence TTAAA, which is probably a target site duplication generated during retrotransposition. A polypurine tract (AAAGGAAAGGAAA) is found adjacent to the 3'-LTR in Pyret and is believed to serve as primer for plus strand cDNA synthesis of the element. The position of the primer binding site (PBS), at which minus strand reverse transcription

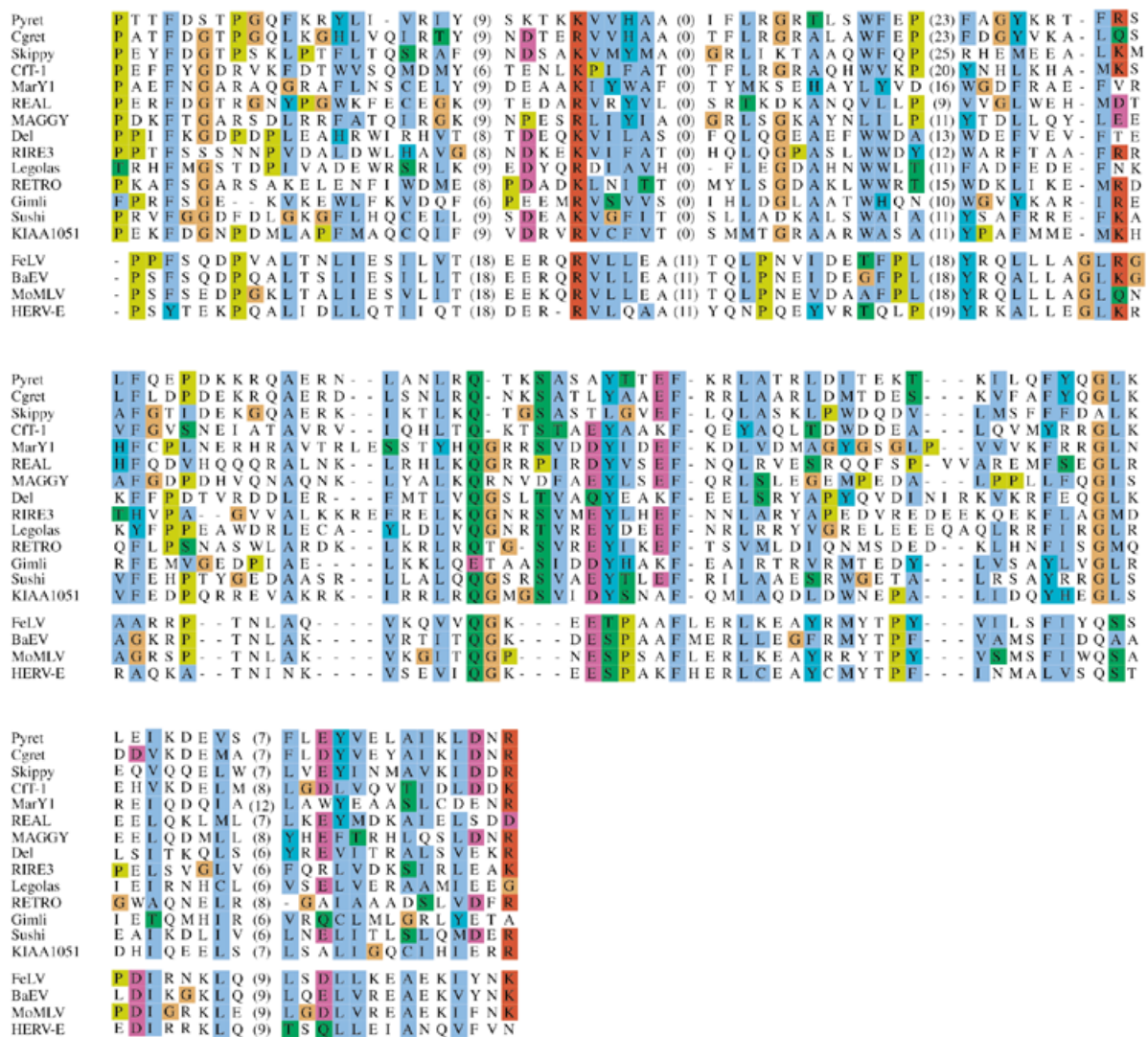


Figure 2. Amino acid alignment of CA-like domains of chromodomain-containing LTR retrotransposons. The bottom block includes authentic CA domains from several gammaretroviruses. Coloring was generated by Clustal X using default values. Accession numbers for the elements in protein databases are as follows: Cgret, AAG24791; Skippy, AAA88790; Cft-1, CAA77890; MarY1, BAA78624; REAL, BAA9271; MAGGY, AAA33419; RETRO, AAD13304; Sushi, AAC33525; KIAA1051, AB028974; Moloney murine leukemia virus (MoMLV), AAB59942; feline leukemia virus (FeLV), AAA43054; baboon endogenous retrovirus (BaEV), BAA00923. Nucleotide accession numbers for the elements which were conceptually translated from their DNA sequences are as follows: Dell, X13886; RIRE3, AB014738; Legolas, AC006570; Gimli, AL049655; human endogenous retrovirus (HERV-E), M10976.

initiates, could not be identified unambiguously in Pyret. A possible PBS candidate is, however, the TGATCGCTCCTG sequence that is found in the 5'-untranslated region. Since the 12 bp sequence CAGGAGAGATCA, which is almost complementary to the PBS candidate, occurs in the middle region of the LTR, a self-primed reverse transcription mechanism proposed for some fungal LTR retrotransposons (29,30), could be adopted in Pyret.

Analysis of hypothetical protein regions

The hypothetical protein product of the ORF of Pyret exhibits significant homology with gene products of known retroelements, especially with those of fungal Ty3/Gypsy retrotransposons. A 14 amino acid sequence that fits the consensus of the zinc finger motif (C-X₂-C-X₃-C) and protease, reverse transcriptase, RNase

H and integrase domains are found in that order, confirming that Pyret belongs to the Ty3/Gypsy superfamily (Metaviridae). This also indicates the presence of a gag-pol ORF in Pyret, which is a rather unique gene structure in LTR retrotransposons. Among sequences available in the databases, the most closely related element to Pyret is Cgret, which was identified in the cranberry fruit rot pathogen *Colletotrichum gloeosporioides* (28). The protease, reverse transcriptase, RNase H and integrase domains of Pyret display 54.5% (55 of 101 residues), 72.5% (124 of 171 residues), 73.2% (93 of 127 residues) and 54.8% (230 of 418 residues) amino acid identity, respectively, to the corresponding domains of Cgret. At the C-terminus of integrase, a chromodomain is observed in Pyret, indicating that Pyret is a member of the chromodomain-containing elements, 'chromoviruses' (3,31).

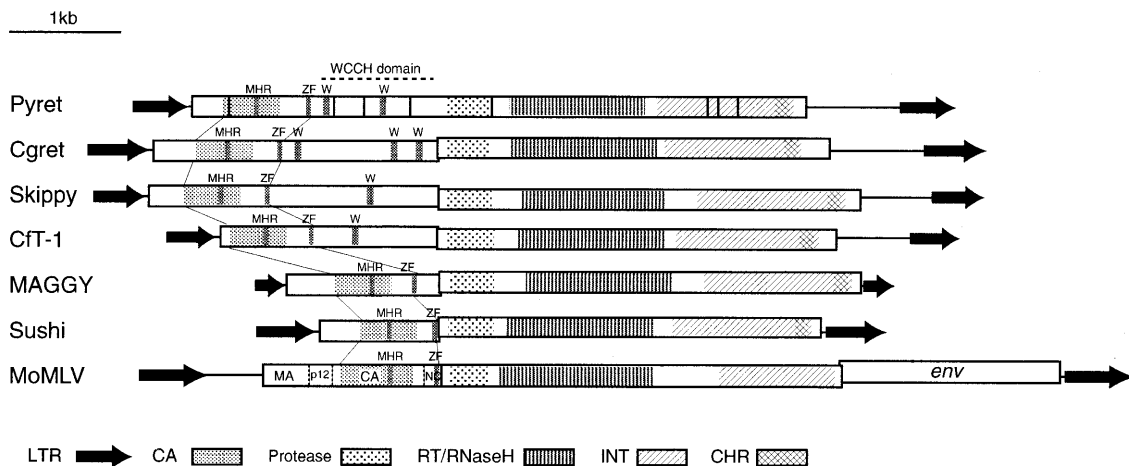


Figure 3. Comparison of overall structure of Pyret with five chromodomain-containing LTR retrotransposons and MoMLV. Boxes indicate ORFs, including putative frames, and vertical lines correspond to stop codons in them. The positions of domains in the gag and pol proteins are indicated. MA, matrix protein; CA, capsid protein; NC, nucleocapsid protein; RT, reverse transcriptase; INT, integrase; CHR, chromodomain; env, envelope protein; MHR, major homology region; ZF, zinc finger motif; W, WCCH motif.

Comparative analysis of the gag protein among 'chromoviruses' and simple retroviruses

The amino acid sequence of the gag gene product is not generally well conserved among distantly related retroviruses (5) and LTR retrotransposons. Therefore, little is known about conserved domains of the gag protein in LTR retrotransposons except for the zinc finger motif, which is conserved in most LTR retrotransposons and retroviruses. In this study, however, multi-alignment of the gag proteins of closely related elements, 'chromoviruses', which include Gypsy/Ty3 LTR retrotransposons and their potential members from fungi, plants and animals, allows us to identify several conserved amino acid sequences in the gag proteins (Fig. 2). We searched for these conserved sequences in the gag proteins of seven clades of 'non-chromoviral' LTR retrotransposons (clades Mdg1, Gypsy, Osvaldo, Mag, Athila, Cer1 and Mdg3), according to Malik and Eickbush (3), and seven genera of retroviruses defined in the most recent report of the ICTV (2). Interestingly, a genus of retroviruses, Gammaretrovirus (mammalian type C retrovirus and MLV-related retrovirus), possesses in its CA domain a greater number of sequences showing recognizable similarity to the conserved sequences, than do other retroviruses examined. Gammaretroviruses are simple retroviruses and include a number of endogenous and exogenous viruses from mammals, birds and reptiles. The type member, Moloney murine leukemia virus (MoMLV), and the other three members, including human endogenous retrovirus E (HERV-E), are presented in the alignment of the gag proteins (Fig. 2).

Referring to the domains defined in MoMLV, the conserved sequences in Figure 2 lie within the CA domain. The CA protein of MoMLV is ~30 kDa, spanning 263 amino acids, and forms the hydrophobic internal core of the viral virion (32,33). The MoMLV sequence corresponding to amino acids 27–217 in the p30 protein (amino acids 243–433 in the MoMLV gag protein) is represented in Figure 2, demonstrating that similarity between MoMLV and chromoviruses can be recognized in most parts of the p30 protein. In all retroviruses, excluding

spumaretroviruses, a stretch of ~20 amino acids in the CA domain is highly conserved and termed the major homology region (MHR) (9). MHR-like sequences are found in most chromoviruses at the corresponding location in the CA domain (Fig. 2). The MHR-like motif in chromoviruses, however, significantly differs from the authentic MHR motif and also shows some sequence variations, including a gap among the chromoviruses. In addition, several members of the chromoviruses, such as grasshopper and Tf1, do not apparently even bear the motif (data not shown). Therefore, the MHR-like motif seems not to be exceptionally conserved in chromoviruses.

The upstream region of the conserved CA domains that compositionally corresponds to the retroviral MA domain or proteins of unknown function (e.g. p12 in MoMLV) is poorly conserved among chromoviruses or barely present in some of the elements (e.g. CFT-1) (Fig. 3). Similarly, little sequence homology exists among the elements downstream of the conserved CA domains except the zinc finger motif. The zinc finger motif occurs twice in the NC domains of most retrovirus groups but once in most gammaretroviruses and not at all in spumaviruses. In most chromoviruses including Pyret the zinc finger motif is present once in the NC domain. The retroviral zinc finger motif is flanked by basic amino acids that have been shown to be important for both RNA packaging and viral infectivity (34,35). Generally, the zinc finger motifs of chromoviruses are also flanked by basic residues, even though the conserved sequence is not clearly identified (data not shown).

Pyret exhibits an extra domain between the NC and protease domains

Structural analysis of Pyret indicates the existence of an additional intervening domain of ~300 amino acids between the NC and protease domains (Fig. 3). The amino acid sequence of the additional domain displays no similarity with known proteins by BLAST search except with the ORF1 products of three closely related fungal retrotransposons, CFT-1, Skippy



Figure 4. Alignment of putative WCCH motifs from four fungal retrotransposons. The bottom block includes WCCH motif-like sequences in various proteins available in the databases. The WCCH motif occurs twice in Pyret (Pyret-1 and Pyret-2) and three times in Cgret (Cgret-1, Cgret-2 and Cgret-3). The numbers after the hyphen are given according to the order of occurrence in the WCCH domain. Coloring was generated by Clustal X using default values. Accession numbers for the proteins in the databases are as follows: *Arabidopsis thaliana* putative protein (*A.thaliana*), CAA16935; tomato leaf curl geminivirus AV2 protein (ToLCV), AAA92806; Lelystad virus orf1A,1B protein (Lelystad), CAA01837.

and Cgret. The length of the additional region varies from ~300 to 450 amino acids depending on the element. Overall homology of the additional domain among the four fungal retrotransposons is generally not high except between Pyret and Cgret. However, a highly conserved sequence spanning 28 amino acids was identified in the additional domains of all the elements. The consensus sequence is H-X₂-h-X-W-X₂-C-X₂-D-X-C-X₂-H-X-(S/T)-X-K-X₄-h-h-P (Fig. 4), where h is a hydrophobic residue. We designate this potential motif as the WCCH motif and, consequently, this additional domain as the WCCH domain. Interestingly, the WCCH motif is present three times in Cgret, twice in Pyret and once in CfT-1 and Skippy (Figs 3 and 4). This repeating nature and the amino acid composition of the consensus indicate that the WCCH motif may be a variation of the zinc finger motif.

We searched for the WCCH motif in other protein sequences in the databases. A small stretch of amino acids similar to the WCCH motif is found in the precoat protein of some geminiviruses (V1 protein or AV2 protein) (Fig. 4). The geminivirus precoat protein is encoded in a different but overlapping ORF with that of coat protein. Although a specific function has not been assigned to the geminivirus precoat protein, its association with viral replication and movement in the host plant has been demonstrated (36). WCCH-like motifs with recognizable similarity were also found in several protein sequences, such as RNA-dependent RNA polymerase of Lelystad virus and a protein of unknown function in the *Arabidopsis* genome. RNA-dependent RNA polymerase and the precoat protein are likely to be nucleic acid-interacting proteins, thereby implying involvement of the WCCH motif in protein–nucleic acid interaction.

Phylogenetic relationship of Pyret to other Gypsy/Ty3 LTR retrotransposons

We performed a phylogenetic analysis using the combined information from three well-conserved domains, reverse transcriptase, RNase H and integrase, according to Malik and Eickbush (3). We focused on fungal Gypsy/Ty3 LTR retrotransposons, including a sushi element from the fish *Fugu rubripes* that is closely related to the fungal elements (37). Only the elements whose sequences are completely or almost completely determined were employed for the analysis. A

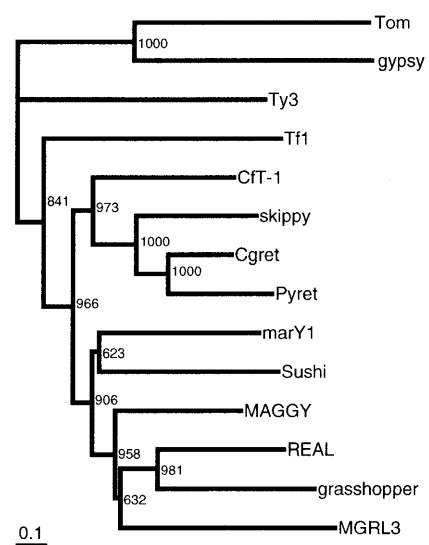


Figure 5. A phylogenetic tree of fungal Ty3/Gypsy LTR retrotransposons based on an alignment of the sum of the amino acids in the reverse transcriptase, RNase H and integrase domains. The tree is constructed by the neighbor-joining method (16) and rooted using insect LTR retrotransposons Gypsy and Tom as outgroups. The number indicates the frequency with which a given branch appeared in 1000 bootstrap replications. Accession numbers for the elements in databases are as follows: Tom, Z24451; gypsy, M12927; Ty3, M34549; Tf1, M38526; CfT-1, Z11866; Skippy, L34658; Cgret, F264028; Pyret, AB062507; marY1, AB028236; Sushi, AF030881; MAGGY, L35053; REAL, AB025309; grasshopper, M77661; MGLR3, AF314096.

phylogenetic tree was constructed based on the neighbor-joining algorithm (16) using two insect Gypsy/Ty3 LTR retrotransposons, Gypsy and Tom, as outgroups (Fig. 5). The tree was subjected to bootstrap resampling of 1000 replicates. The WCCH domain-containing elements, including CfT-1, Skippy, Cgret and Pyret, formed a monophyletic group that stands apart from another fungal group that includes WCCH domain-less elements, MAGGY, grasshopper, REAL, MGLR-3 and marY1, and the fish element sushi. This indicates that Gypsy/Ty3 LTR retrotransposons from filamentous fungi and fish could be divided into two groups with respect to the presence or absence of the WCCH domain based on the pol sequence. Two yeast elements, Ty3 and Tf1, were more distantly related to the other fungal elements and the fish element as reported in previous papers (3,31,38).

Distribution of Pyret-related elements in *Pyricularia* isolates

Finally, the distribution of Pyret-related elements in *Pyricularia* spp. was examined. We use the genus name of the anamorph, *Pyricularia*, in this section since the teleomorph has not been determined in some *Pyricularia* species and isolates. *Pyricularia* spp. include several host-specific pathotypes that can infect only a limited range of host plant species, e.g. the *Oryza* (rice-infecting isolates), *Triticum* (wheat-infecting isolates) and *Digitaria* (crabgrass-infecting isolates) pathotypes (39). Those isolates are genetically diverse with respect to rDNA, nuclear DNA and distribution of transposons (11,39–41). A Southern blot with 12 *Pyricularia* isolates from 11 plant species is given in Figure 6. In contrast to the previously

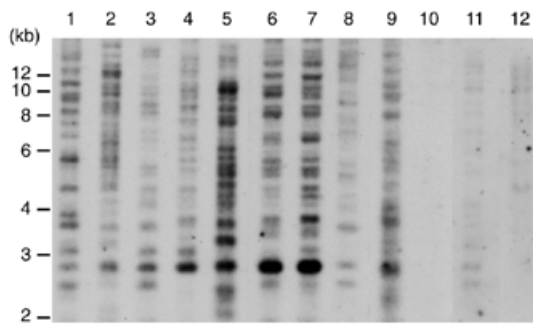


Figure 6. DNA gel blot analysis of the distribution of Pyret-related elements in *Pyricularia* isolates. Genomic DNA was digested with *EcoRI* and probed with a Pyret probe (see Materials and Methods). Isolates used and their host plants were as follows: lane 1, UG-77-7-1-1 (*Eleusine indica*); lane 2, G10-1 (*E. coracana*); lane 3, NI986 (*Eragrostis lehmanniana*); lane 4, Br58 (*Avena sativa*); lane 5, 0903-4 (*Oryza sativa*); lane 6, Br127.11 (*Triticum aestivum*); lane 7, Br130.1 (*T. aestivum*); lane 8, NI919 (*Leersia oryzoides*); lane 9, NI979 (*Pennisetum clandestinum*); lane 10, HYZIM101-1-1-1 (*Zingiber mioga*); lane 11, NI907 (*Digitaria sanguinalis*); lane 12, Br29 (*Digitaria horizontalis*).

reported *Pyricularia* LTR retrotransposons MAGGY and grasshopper, which are distributed only in specific pathotypes of the fungus (25,42,43), Pyret-related elements exhibit a wide ranging distribution pattern in the *Pyricularia* population. Actually, Pyret-related elements could be detected by Southern analysis in 75 of 79 *Pyricularia* isolates from 31 monocot plant species (data not shown), suggesting that Pyret may be an 'old' element that was acquired in the *Pyricularia* population at an early stage of its differentiation.

DISCUSSION

We have isolated a novel LTR retrotransposon, Pyret, in the causal pathogen of blast fungus, *M. grisea*. Pyret is unique because it exhibits an extra domain (WCCH domain) between the NC and protease domains in addition to the common components of retroviruses and LTR retrotransposons. The WCCH domain contains one to three copies of the WCCH motif whose consensus is H-X₂-h-X-W-X₂-C-X₂-D-X-C-X₂-H-X-(S/T)-X-K-X₄-h-h-P (where h is a hydrophobic residue). Some retroviruses, exemplified by mouse mammary tumor virus (MMTV), also show an extra domain between the NC and protease domains and actually produce a corresponding gag-pro transframe protein (p30) possessing both zinc-binding and dUTPase activity (44,45). However, p30 and a putative protein encoded by the WCCH domain seem to be functionally different, for the following reasons: (i) p30 does not contain the WCCH motif; (ii) the consensus sequence of dUTPase is not found in any of the WCCH domains encoded by four fungal elements; (iii) the WCCH domain is within ORF1 while p30 is translated in a gag-pro transframe manner. Thus, the function of the WCCH domain is unclear to date.

One surprising finding in this study is that Pyret, as well as other members of the 'chromoviruses', a group of Gypsy/Ty3 LTR retrotransposons, displays significant sequence homology in the CA domain of the gag protein with most members of Gammaretrovirus, a genus of simple retroviruses.

Even among retroviruses, little sequence homology is known in the gag protein except for the MHR and the zinc finger motif when they belong to distantly related retrovirus groups (5). Another common feature between gammaretroviruses and 'chromoviruses' is that the zinc finger motif is present once in the NC domains of most of the members while it occurs twice in most retroviruses. Furthermore, Malik and Eickbush (3) reported that the GPY/F module, which is well-conserved in the integrase domains of Gypsy/Ty3 LTR retrotransposons, is also found in the integrase of some Gammaretrovirus members. These facts suggest that gammaretroviruses are relatively close to Gypsy/Ty3 LTR retrotransposons, especially to chromoviruses. At least with regard to the CA domain in the gag protein, chromoviruses and gammaretroviruses may have roots in a common ancestor. Alternatively, they may retain the features of the common prototype of Retroelementopsida, while the other members have modified their sequences and/or acquired other components during diversification and evolution.

In *Pyricularia*, four Gypsy/Ty3 LTR retrotransposons have been identified so far, grasshopper, MAGGY, MGLR-3 and Pyret (25,26,42). The former two elements are distributed in a pathotype-specific manner while the latter two are more ubiquitously distributed (25,26,42,43). Among these, MAGGY is the only element that has been shown to have transpositional activity to date (12,46). We assessed the activity of Pyret by comparing the fingerprinting patterns between sister progeny obtained from a cross between foxtail millet and wheat isolates of *P. oryzae* as described by Eto *et al.* (11). Many DNA rearrangements near transposable elements were reported between sister progeny in Southern blots probed with MAGGY, MGR586 and Pot2, indicating that these elements are active (11). However, few rearrangements of Pyret bands were observed in 40 pairs of sister progeny examined (unpublished data), suggesting that Pyret is not a very active element.

In a BLAST search a 0.3 kb region in Pyret (nt 1649–1986) was found to be 91.1% identical to MGSR1 (*M. grisea* short repeat 1), which is reported as a SINE-like repetitive sequence in *M. grisea* (47). This putative element is 336 bp long, delimited by 8 bp direct repeats, and was cloned from a 0.8 kb *XhoI* fragment in a rice isolate of *M. grisea*. The 0.8 kb *XhoI* fragment was commonly found in six λ phage clones from different loci. Thus, the 0.8 kb *XhoI* fragment itself seems to be repetitive. A corresponding region to the 0.8 kb *XhoI* fragment was identified in the NC and WCCH domains of our Pyret clone and also in those of Cgret with recognizable similarity. These data suggest that MGSR1 may be a conceptual element, even though we cannot rule out the possibility that a SINE-like element might have transposed into the prototype of Pyret in ancient times.

Even though the WCCH domain identified in this study was found in only four fungal Gypsy/Ty3 LTR retrotransposons, WCCH-like motifs were recognized not only in several viral proteins but also in proteins of unknown function in the human and *Arabidopsis* genomes (data not shown). Therefore, this putative module may have some function in a wide range of organisms. Experimental evidence of RNA- and/or DNA-binding ability of the WCCH motif is of interest and will be addressed in the future.

ACKNOWLEDGEMENTS

We thank Dr Sally Leong for providing the cosmid vector pMLF2 and Dr Laura Darnielle for critical reading of the manuscript. This work was partly supported by a grant from the Ministry of Education, Science, Sports and Culture of Japan (no. 13660050).

REFERENCES

- Hull, R. (1999) Classification of reverse transcribing elements: a discussion document. *Arch. Virol.*, **144**, 209–214.
- Pringle, C. (1999) Virus taxonomy—1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Arch. Virol.*, **144**, 421–429.
- Malik, H.S. and Eickbush, T.H. (1999) Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J. Virol.*, **73**, 5186–5190.
- Xiong, Y. and Eickbush, T.H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.*, **9**, 3353–3362.
- Wills, J.W. and Craven, R.C. (1991) Form, function and use of retroviral Gag proteins. *AIDS*, **5**, 639–654.
- Stromberg, K., Hurley, N., Davis, N., Rueckert, R. and Fleissner, E. (1974) Structural studies of avian myeloblastosis virus: comparison of polypeptides in virion and core component by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Virol.*, **13**, 513–528.
- Gorelick, R., Henderson, L., Hanser, J. and Rein, A. (1988) Point mutants of Moloney murine leukemia virus that fail to package viral RNA: evidence for specific RNA recognition by a “zinc finger-like” protein sequence. *Proc. Natl Acad. Sci. USA*, **85**, 8420–8424.
- Berkowitz, R., Fisher, J. and Goff, S. (1996) RNA packaging. *Curr. Top. Microbiol. Immunol.*, **214**, 177–218.
- Patarca, R. and Haseltine, W. (1985) A major retroviral core protein related to EPA and TIMP. *Nature*, **318**, 390.
- Orlinsky, K.J., Gu, J., Hoyt, M., Sandmeyer, S. and Menees, T.M. (1996) Mutations in the Ty3 major homology region affect multiple steps in Ty3 retrotransposition. *J. Virol.*, **70**, 3440–3448.
- Eto, Y., Ikeda, K., Chuma, L., Kataoka, T., Kuroda, S., Kikuchi, N., Don, L.D., Kusaba, M., Nakayashiki, H., Tosa, Y. and Mayama, S. (2001) Comparative analyses of the distribution of various transposable elements in *Pyricularia* and their activity during and after the sexual cycle. *Mol. Gen. Genet.*, **264**, 565–577.
- Nakayashiki, H., Kiyotomi, K., Tosa, Y. and Mayama, S. (1999) Transposition of the retrotransposon MAGGY in heterologous species of filamentous fungi. *Genetics*, **153**, 693–703.
- Altschul, S., Madden, T., Schaffer, A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
- Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
- Henikoff, S. and Henikoff, J.G. (1992) Amino acid substitution matrices for protein blocks. *Proc. Natl Acad. Sci. USA*, **89**, 10915–10919.
- Saito, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
- Page, R. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.*, **12**, 357–358.
- Nishimura, M., Nakamura, S., Hayashi, N., Asakawa, S., Shimizu, N., Kaku, H., Hasebe, A. and Kawasaki, S. (1998) Construction of a BAC library of the rice blast fungus *Magnaporthe grisea* and finding specific genome regions in which its transposons tend to cluster. *Biosci. Biotechnol. Biochem.*, **62**, 1515–1521.
- Nitta, N., Farman, M.L. and Leong, S.A. (1997) Genome organization of *Magnaporthe grisea*: integration of genetic maps, clustering of rearrangeable elements and identification of genome duplications and rearrangements. *Theor. Appl. Genet.*, **95**, 20–32.
- Romao, J. and Hamer, J.E. (1992) Genetic organization of a repeated DNA sequence family in the rice blast fungus. *Proc. Natl Acad. Sci. USA*, **89**, 5316–5320.
- Cambareri, E.B., Jensen, B.C., Schabtach, E. and Selker, E.U. (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. *Science*, **244**, 1571–1575.
- Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y. and Mayama, S. (1999) Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Mol. Gen. Genet.*, **261**, 958–966.
- Anaya, N. and Roncero, M.I. (1995) Skippy, a retrotransposon from the fungal plant pathogen *Fusarium oxysporum*. *Mol. Gen. Genet.*, **249**, 637–647.
- Dioliz, A., Marches, F., Fortini, D. and Brygoo, Y. (1995) Boty, a long-terminal-repeat retroelement in the phytopathogenic fungus *Botrytis cinerea*. *Appl. Environ. Microbiol.*, **61**, 103–108.
- Dobinson, K.F., Harris, R.E. and Hamer, J.E. (1993) Grasshopper, a long terminal repeat (LTR) retroelement in the phytopathogenic fungus *Magnaporthe grisea*. *Mol. Plant Microbe Interact.*, **6**, 114–126.
- Kang, S. (2001) Organization and distribution pattern of MGLR-3, a novel retrotransposon in the rice blast fungus *Magnaporthe grisea*. *Fungal Genet. Biol.*, **32**, 11–19.
- McHale, M.T., Roberts, I.N., Noble, S.M., Beaumont, C., Whitehead, M.P., Seth, D. and Oliver, R.P. (1992) CFT-I: an LTR-retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato. *Mol. Gen. Genet.*, **233**, 337–347.
- Zhu, P. and Oudemans, P.V. (2000) A long terminal repeat retrotransposon Cgret from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. *Curr. Genet.*, **38**, 241–247.
- Levin, H. (1995) A novel mechanism of self-primed reverse transcription defines a new family of retroelements. *Mol. Cell. Biol.*, **15**, 3310–3317.
- Lin, J.H. and Levin, H. (1997) Self-primed reverse transcription is a mechanism shared by several LTR-containing retrotransposons. *RNA*, **3**, 952–953.
- Marin, I. and Llorens, C. (2000) Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. *Mol. Biol. Evol.*, **17**, 1040–1049.
- Naso, R., Arcement, L. and Arlinghaus, R. (1975) Biosynthesis of Rauscher leukemia viral proteins. *Cell*, **4**, 31–36.
- Naso, R., Karshin, W., Wu, Y. and Arlinghaus, R. (1979) Characterization of 40,000- and 25,000-dalton intermediate precursors to Rauscher murine leukemia virus gag gene products. *J. Virol.*, **32**, 187–198.
- Bowzard, J., Bennett, R., Krishna, N., Ernst, S., Rein, A. and Wills, J. (1998) Importance of basic residues in the nucleocapsid sequence for retrovirus Gag assembly and complementation rescue. *J. Virol.*, **72**, 9034–9044.
- Housset, V., De Rocquigny, H., Roques, B. and Darlix, J. (1993) Basic amino acids flanking the zinc finger of Moloney murine leukemia virus nucleocapsid protein NCp10 are critical for virus infectivity. *J. Virol.*, **67**, 2537–2545.
- Padidam, M., Beachy, R. and Fauquet, C. (1996) The role of AV2 (“precoat”) and coat protein in viral replication and movement in tomato leaf curl geminivirus. *Virology*, **224**, 390–404.
- Poulter, R. and Butler, M. (1998) A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. *Gene*, **215**, 241–249.
- Kaneko, I., Tanaka, A. and Tsuge, T. (2000) REAL, an LTR retrotransposon from the plant pathogenic fungus *Alternaria alternata*. *Mol. Gen. Genet.*, **263**, 625–634.
- Kato, H., Yamamoto, M., Yamaguchi-Ozaki, T., Kadouchi, H., Iwamoto, Y., Nakayashiki, H., Tosa, Y., Mayama, S. and Mori, N. (2000) Pathogenicity, mating ability and DNA restriction fragment length polymorphisms of *Pyricularia* populations isolated from *Gramineae*, *Bambusideae* and *Zingiberaceae* plants. *J. Gen. Plant Pathol.*, **66**, 30–47.
- Hamer, J., Farrall, L., Orbach, M., Valent, B. and Chumley, F. (1989) Host species-specific conservation of a family of repeated DNA sequences in the genome of a fungal plant pathogen. *Proc. Natl Acad. Sci. USA*, **86**, 9981–9985.
- Kusaba, M., Eto, E., Tosa, Y., Nakayashiki, H. and Mayama, M. (1999) Genetic diversity in *Pyricularia* isolates from various hosts revealed by polymorphisms of nuclear ribosomal DNA and the distribution of the MAGGY retrotransposon. *Ann. Phytopathol. Soc. Jpn.*, **65**, 588–596.
- Farman, M.L., Tosa, Y., Nitta, N. and Leong, S.A. (1996) MAGGY, a retrotransposon in the genome of the rice blast fungus *Magnaporthe grisea*. *Mol. Gen. Genet.*, **251**, 665–674.
- Tosa, Y., Nakayashiki, H., Hyodo, H., Mayama, S., Kato, H. and Leong, S.A. (1995) Distribution of retrotransposon MAGGY in *Pyricularia* species. *Ann. Phytopathol. Soc. Jpn.*, **61**, 549–554.

44. Bergman,A., Bjornberg,O., Nord,J., Nyman,P. and Rosengren,A. (1994) The protein p30, encoded at the gag-pro junction of mouse mammary tumor virus, is a dUTPase fused with a nucleocapsid protein. *Virology*, **204**, 420–424.
45. Koppe,B., Menendez-Arias,L. and Oroszlan,S. (1994) Expression and purification of the mouse mammary tumor virus gag-pro transframe protein p30 and characterization of its dUTPase activity. *J. Virol.*, **68**, 2313–2319.
46. Shull,V. and Hamer,J. (1996) Rearrangements at a DNA-fingerprint locus in the rice blast fungus. *Curr. Genet.*, **30**, 263–271.
47. Sone,T., Suto,M. and Tomita,F. (1993) Host species-specific repetitive DNA sequence in the genome of *Magnaporthe grisea*, the rice blast fungus. *Biosci. Biotechnol. Biochem.*, **57**, 1228–1230.