CsRn1, a Novel Active Retrotransposon in a Parasitic Trematode, *Clonorchis sinensis*, Discloses a New Phylogenetic Clade of Ty3/gypsy-like LTR Retrotransposons

Young-An Bae,* Seo-Yun Moon,* Yoon Kong,† Seung-Yull Cho,† and Mun-Gan Rhyu*

*Department of Microbiology, College of Medicine, Catholic University of Korea, Seoul, Korea; and †Department of Molecular Parasitology, Sungkyunkwan University School of Medicine, Suwon, Korea

We screened the genome of a trematode, *Clonorchis sinensis*, in order to identify novel retrotransposons and thereby provide additional information on retrotransposons for comprehensive phylogenetic study. Considering the vast potential of retrotransposons to generate genetically variable regions among individual genomes, randomly amplified polymorphic DNAs (RAPDs) detected by arbitrarily primed polymerase chain reactions were selected as candidates for retrotransposon-related sequences. From RAPD analysis, we isolated and characterized a novel retrotransposon in *C. sinensis* as the first member of uncorrupted long-terminal-repeat (LTR) retrotransposons in phylum Platyhelminthes. The retrotransposon, which was named *Clonorchis sinensis Retrotransposon 1* (CsRn1), showed a genomewide distribution and had a copy number of more than 100 per haploid genome. CsRn1 encoded an uninterrupted open reading frame (ORF) of 1,304 amino acids, and the deduced ORF exhibited similarities to the pol proteins of Ty3/gypsy-like LTR retrotransposons. The mobile activity of master copies was predicted by sequence analysis and confirmed by the presence of mRNA transcripts. Phylogenetic analysis of Ty3/gypsy-like LTR retrotransposons detected a new clade comprising CsRn1, *Kabuki* of *Bombyx mori*, and an uncharacterized element of *Drosophila melanogaster*. With its high repetitiveness and preserved mobile activity, it is proposed that CsRn1 may play a significant role in the genomic evolution of *C. sinensis*.

Introduction

A remarkable wealth of data is now available on the diversity and distribution of retrotransposons in nearly all eukaryotes (see Boeke and Stoye 1997 and references therein). As a class of transposable elements, retrotransposons move and integrate into new sites within genomes via reverse transcription of an RNA intermediate (Boeke et al. 1985) and can therefore provide genetic variations ranging from simple sequence polymorphism within the elements to dramatic alterations in chromosomal structure (Kidwell and Lisch 1997; Fedoroff 2000). Based on such vast potential to generate genetic variations, retrotransposons have been known as major agents in evolution that give rise to phenotypic variants (Long et al. 2000) and, in the long term, drive speciation (McDonald 1990).

With the cumulative data on retrotransposons, various studies have discussed the probable evolutionary course of retrotransposons, including studies of long-terminal-repeat (LTR) family, non-LTR family, and exogenous retroviruses (Xiong and Eickbush 1990; Malik, Burke, and Eickbush 1999; Malik and Eickbush 1999). However, the phylogenetic relationship of diverse retrotransposons remains ambiguous, mainly with regard to factors concerning the branching patterns in phylegetic trees, such as the formation of polymy (Malik and Eickbush 1999) and polyphyletic distribution (Malik and Eickbush 1999; Marin and Llórens 2000). Thus, the number of ancient classes responsible for diverse clades of retrotransposons and the extent of possible horizontal transfer between different species are difficult to define. Such ambiguities are likely to arise from limitations in sequence information on retrotransposons and unbalanced sampling from each taxon.

There have been a number of reports on the LTR retrotransposons of animals such as nematodes (Felder et al. 1994; Bowen and McDonald 1999), insects (Lindsley and Zimm 1992; Biessmann et al. 1999; Abe et al. 2000), echinoderms (Britten et al. 1995), and fish (Poulter and Butler 1998). In Platyhelminthes, however, *Gulliver of Schistosoma japonicum* (Laha et al. 2001) is the only full-length LTR retrotransposon so far described in the phylum, although Arkhipova and Meselson (2000) have recently reported a segmental sequence of an LTR retrotransposon in *Dugesia*. Moreover, the sequence of *Gulliver* is corrupted, even though its expression has been demonstrated at the level of transcription by reverse transcription-PCR (RT-PCR) (Laha et al. 2001). Thus, currently there are no reports on uncorrupted LTR retrotransposons, which may play a significant role in the formation of genomes in Platyhelminthes, including trematodes.

Retrotransposons introduce variations through their heterogeneous integration and subsequent sequence divergence, and these polymorphic regions can be identified as randomly amplified polymorphic DNAs (RAPDs) by arbitrarily primed PCR (AP-PCR) (Abe et al. 1998). In the present study, we attempted to identify retrotransposons from the genome of *Clonorchis sinensis*, an important human liver fluke in East Asia, based on the analysis of RAPD sequences. By screening variable genetic regions from individual *C. sinensis* using the AP-PCR method, a retrotransposon, named *Clonorchis sinensis Retrotransposon 1* (CsRn1), was isolated.

Abbreviations: IN, integrase; PBS, primer-binding site; PPT, polypurine tract; PR, protease; RT, reverse transcriptase; TSD, target site duplication; RH, RNase H.

Key words: *Clonorchis sinensis*, trematode, LTR retrotransposon, AP-PCR, RAPD, master copy.

Address for correspondence and reprints: Mun-Gan Rhyu, Department of Microbiology, College of Medicine, Catholic University of Korea, Seoul 137-701, Korea. E-mail: rhyumung@cmc.cuk.ac.kr.
as the second complete but the first uncorrupted LTR retrotransposon identified in the phylum Platyhelminthes. Structural and genomic analyses of CsRn1 and its phylogenetic relationship with other Ty3/gypsy-like LTR retrotransposons are presented.

Materials and Methods

Parasite and DNA Extraction

Adult C. sinensis were collected from the livers of experimental rabbits which were challenged orally with metacercariae obtained from naturally infected fish in Kimhae, Korea, 3 months prior to dissection. The worms were washed with physiological saline five times at 4°C. Fresh worms were used immediately for DNA extraction with the Wizard DNA Purification Kit (Promega, Madison, Wis.) according to the manufacturer’s instructions.

AP-PCR and Cloning of Individual Worm-Specific Products

Genomic DNAs extracted from individual worms were used for the amplification of worm-specific RAPD regions by PCR under low-stringency conditions. The following arbitrarily designed primers were used in the PCR reactions: AP-1 (5'-GATCCGGTTCA-3'), AP-3 (5'-ACCCATAACC-3'), B4 (5'-GGACTTGAGT-3'), B5 (5'-TGGCGCCCTTC-3'), B8 (5'-TCCACACGG-3'), B12 (5'-CTTGAGCGA-3'), B13 (5'-TCTCCCCCCT-3'), F6 (5'-GGGAAATTGG-3'), and F15 (5'-CCAGTACTCC-3'). The reaction mixtures included 40 ng of genomic DNA, 1.25 µM of primers, 0.2 mM each of dATP, dGTP, dCTP, and dTTP, and 1.25 U of Taq polymerase (Takara, Shiga, Japan) in a total reaction volume of 20 µl. PCR conditions were as follows: 32 cycles of 1 min at 94°C, 1 min at 37°C, and 2 min at 72°C, and a final extension of 10 min at 72°C. The reproducibility of the results was tested by repeating the reactions under identical conditions three times. The PCR products were fractionated by electrophoresis on agarose gels and stained with ethidium bromide. Individual-specific bands were recovered from agarose gels, reamplified with the corresponding primers, and then cloned into pGEM-T Easy Vector (Promega) for nucleotide sequencing.

Southern Blot Hybridization

Five micrograms of genomic DNAs isolated from C. sinensis adult worms were digested with restriction enzymes. After being fractionated through 0.8% agarose gels, the DNAs were blotted onto nylon membrane (Hybond-N+; Amersham Pharmacia Biotech, Uppsala, Sweden) by capillary action in 10 × standard saline citrate (SSC) at 94°C, 1 min at 37°C, and 2 min at 72°C. The membranes were washed twice in 6 M urea, 0.4% sodium dodecyl sulfate, and 0.1 × SSC at 42°C for 20 min and twice in 2 × SSC at room temperature for 5 min.

Dot-Blot Analysis

Ten micrograms of genomic DNA were blotted onto nylon membrane according to the standard procedure of dot-blot analysis (Sambrook, Fritsch, and Maniatis 1989). Two identical membranes were prepared, of which one was hybridized with probe for repetitive sequences and the other was hybridized with that for cysteine protease as a single copy control (GenBank accession number AF271091). The probe for protease was amplified from C. sinensis genomic DNA with the CsCP3-S1 (5'-GCTTGGACTCCGACTACCCATAG-3') and CsCP3-R3 (5'-GGTTTAAACGATTGTCATCCG-3') primers. Probe labeling and hybridizing conditions were same as those for Southern blot hybridization. The intensities of signals were measured using the LAS-1000plus system (FUJIFILM, Tokyo, Japan).

Construction and Screening of Genomic DNA Library

DNAs from adult worms were partially digested with Sau3AI (Takara). DNA fragments of 9–23 kb were recovered by ultracentrifugation onto sucrose gradients, purified, and then cloned into lambda FIX II vector predigested with XhoI (Stratagene, La Jolla, Calif.). The constructs were packaged into lambda particles using Gigapack III Gold-11 packaging extract. The unamplified libraries were screened with DNA probe labeled with the ECL labeling kit. The conditions for plaque-lift hybridization were identical to those for Southern blot hybridization. The inserts of lambda clones were amplified by long PCR (Chen, Fockler, and Higuchi 1994) using primers designed from vector regions (5'-CTAATACGACTCACTATAGGGCGTCG-3' and 5'-CCCTCACAAGGAGTGAAGGTCGAC-3') and LA Taq polymerase according to the standard cycle conditions (Takara). The amplified products were digested with restriction enzymes and were then cloned into pBluescript II SK(−) phagemid (Stratagene) for nucleotide sequencing.

Screening of cDNA Library by PCR Methods

A cDNA library of adult C. sinensis was constructed in λZAP II vector using a cDNA synthesis kit (Stratagene) according to the manufacturer’s instructions. The library was amplified and was then used in standard PCR reactions for the detection of mRNA transcripts. The PCR products were cloned into pGEM-T Easy Vector for sequencing.

Sequence Analysis

The nucleotide sequences were automatically determined with an ABI PRISM 377 DNA sequencer (Applied Biosystems, Foster City, Calif.) and a BigDye Terminator Cycle Sequencing Reaction Kit (Perkin Elmer Corporation, Foster City, Calif.). To ensure the accuracy of sequencing reactions, sequences of single strands
from five clones of vector-ligated DNA fragments were determined. For PCR products obtained during cDNA library screening, nucleotide sequences from both strands were determined. After sequencing, homology searches were performed against the nonredundant database at the National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/) using BLASTN and BLASTX (Altschul et al. 1997). The REPEAT program in the GCG package, version 8. (University of Wisconsin), was used to determine the direct repeat sequences. The putative open reading frames (ORFs) were predicted by GeneScan (Burge and Karlin 1997), GeneMark (Borodovsky and Lukashin, unpublished; http://genemark.biology.gatech.edu/GeneMark/), and ORF Finder (NCBI). A search for the functional protein domains was performed using ProDom (Gouzy, Corpet, and Kahn 1999) and ProfileScan (Gribskov, McLachlan, and Eisenberg 1987).

Phylogenetic Analysis

The nucleotide sequences were aligned with ClustalX (Thompson et al. 1997). After optimizing the sequence alignments using the PHYDIT program (Chun 1995), divergence values were calculated and a dendrogram was drawn using the programs DNADIST and NDJEBIOR, respectively, of PHYLIP (Felsenstein 1993). Based on the previous reports (Xiong and Eickbush 1990; Malik and Eickbush 1999), reverse transcriptase (RT), RNase H (RH), and integrase (IN) domains in pol proteins were defined from alignments using CLUSTAL W (Thompson, Higgins, and Gibson 1994), and amino acid sequences of the three domains were combined for the phylogenetic analysis of LTR retrotransposons. After aligning the combined sequences with ClustalX and optimizing the alignment with GeneDoc (Nicholas and Nicholas 1997), a phylogenetic analysis was performed using PROTDIST and NDJEBIOR of PHYLIP. The trees were displayed by TreeView (Page 1996), and the statistical significance of branching points was evaluated with 1,000 random samplings of the input sequence alignments using SEQBOOT.

Results

Isolation of RAPDs by AP-PCR

With nine 10mer arbitrary primers, singly or in pairs, AP-PCRs were performed to find band shifts in genomic DNAs separately extracted from individual C. sinensis worms. In AP-PCRs with DNAs from three worms, three individual-specific band shifts were discovered (fig. 1A), and their sequences were used to search the GenBank database using the BLAST algorithm. The sequence of a band shift with the B13 primer, named P-B13 (GenBank accession number AZ551682), showed significant identity at the amino acid (aa) sequence level to pol proteins in Kabuki of Bombyx mori (AB032718; identity value of 34%) and in bovine syncytial virus (U94514, 23%). Significant identities to any known genes were not found in the homology searches.

Fig. 1.—Arbitrarily primed-PCR (AP-PCR) analysis of individual worms for isolating randomly amplified polymorphic DNAs (RAPD). A, Electrophoresis of AP-PCR products on agarose gel visualized by ethidium bromide staining. Separately extracted genomic DNAs from individual Clonorchis sinensis worms were amplified using AP primers. Band shifts corresponding to individual-specific RAPD markers are indicated by arrows. Numbers at the top refer to individual worms, and primers used in each reaction are indicated above the numbers. The positions of DNA size standards (in bp) are shown on the left.

B, Southern blot hybridization of P-B13 to genomic DNAs of C. sinensis. Restriction endonucleases used for the digestion of DNAs are indicated at the top. The positions of DNA size standards (in kb) are shown on the right.
using the sequences of the other two bands. To determine the genomic distribution of P-B13, the P-B13 DNA fragment was used to probe a Southern blot of genomic DNA from *C. sinensis*. Multiple bands of hybridization to the genomic DNA digested with a series of restriction enzymes indicated that P-B13 represents a bridization to the genomic DNA. Duplicated target sites of 4 bp are represented as target site duplications (TSDs).

Striped boxes within the open reading frame (ORF) show the conserved functional domains of Gag, protease (PR), reverse transcriptase (RT), RNase H (RH), and three subdomains of integrase (IN). Duplicated target sites of 4 bp are represented as target site duplications (TSDs).

The overall structure of a novel retrotransposon encompassing the P-B13 marker was determined from three randomly selected lambda clones, labeled CsRn1, Cs-2, and Cs-4, obtained by screening the genomic DNA library with the P-B13 probe (fig. 2). In each of these three clones, the full lengths of the CsRn1 retrotransposon differed slightly due to several insertions and deletions (indels), and a consensus sequence determined from the three clones had a size of 5,026 bp. The whole units of CsRn1 were bounded by direct repeats of 4 bp known as target site duplications (TSDs) introduced during the process of integration.

The complete CsRn1 elements were flanked by LTRs with an average size of 471 bp (fig. 2). As in most LTR retrotransposons (Boeke and Stoye 1997), the LTRs contained at both ends short inverted repeats of 3 bp that initiated as TG and terminated as CA (5'-TGT . . . ACA-3'). A sequence motif (AATACA) similar to a typical poly A signal sequence (AATAAA) was found within the LTR sequence, but a promoter signal sequence (TATA box) could not be defined. The sequence of 12 bp adjacent to the 5'-LTR was complementary to the 3'-end of bovine tRNA^Trp_. Thus, this region seems to be a likely primer-binding site (PBS) for the synthesis of the first cDNA strand. An additional priming site (5'-GGGGGAGTAG-3') for the synthesis of the second cDNA strand (poly-purine tract [PPT]) was found in the direct upstream region of 3'-LTR (fig. 3).

An internal region of the CsRn1 copy from lambda clone 4 (λCs-4) contained one large, uninterrupted ORF (1,304 aa). The copies from the other two lambda clones (λCs-1 and λCs-2) showed similar results after the sequences were corrected for corruptions. The deduced ORF included well-conserved functional domains in the order protease (PR), RT, RH, and IN (fig. 2). Instead of a conventional Gag motif (CCCH), the ORF had an apparent nucleic-acid-binding site (CHCC) at the predicted 3' ends of Gag just prior to the DTG aspartic protease active site. Although the CHCC motif is unusual, it is also repetitively found in Gag proteins of *Kabuki* and AE003787 (fig. 4). The amino acid conservation (fig. 4) and domain order (PR-RT-RH-IN) in the ORF indicated that CsRn1 belongs to a family of Ty3/gypsy-like LTR retrotransposons (Pringle 1999).

Identification of CsRn1 Master Copies

CsRn1 elements were interspersed throughout the genome of *C. sinensis*, rather than being tandemly arrayed or clustered at limited loci, and the copy number was estimated to be approximately 100 per haploid genome under high-stringency conditions (0.1 × SSC) (fig. 5). In addition to the 3 initially identified copies, 11 copies of full-length CsRn1 elements were further obtained from the *C. sinensis* genomic library for comparison of their overall sequences and structures. The 14 CsRn1 copies (GenBank accession numbers AY013558–AY013571) shared an overall sequence identity of 98%.

![Structure of a Novel LTR Retrotransposon, CsRn1](image)

**Fig. 2.—Overall structure of the CsRn1 long-terminal-repeat (LTR) retrotransposon.** Boxes containing black arrows represent flanking LTRs. Striped boxes within the open reading frame (ORF) show the conserved functional domains of Gag, protease (PR), reverse transcriptase (RT), RNase H (RH), and three subdomains of integrase (IN). Duplicated target sites of 4 bp are represented as target site duplications (TSDs). Positions of P-B13 RAPD and LTR probe for southern hybridization (P-LTR) are indicated. A restriction map of the full-length genomic DNA fragment was used to probe a Southern blot of genomic DNA digested with a series of restriction enzymes indicated that P-B13 represents a bridization to the genomic DNA. Duplicated target sites of 4 bp are represented as target site duplications (TSDs). Positions of P-B13 RAPD and LTR probe for southern hybridization (P-LTR) are indicated. A restriction map of the full-length genomic DNA fragment was used to probe a Southern blot of genomic DNA digested with a series of restriction enzymes indicated that P-B13 represents a bridization to the genomic DNA. Duplicated target sites of 4 bp are represented as target site duplications (TSDs).

**Fig. 3.—Sequences of the putative primer-binding site (PBS) and polypurine tract (PPT) of CsRn1.** The sequences of 5' long-terminal-repeat (LTR) termini, PBS, PPT, and 3' LTR termini are aligned and presented. The 3' end of bovine tRNA^Trp_ is also presented in its corresponding region. Dots represent gaps introduced into sequences to increase their similarity. Sequences of Kabuki of *Bombyx mori* (AB032718) and an undescribed LTR retrotransposon of *Drosophila melanogaster* on AE003787 are used for the alignment.
Fig. 4.—Multiple-sequence alignments of functional protein domains from various long-terminal-repeat (LTR) retrotransposons. The amino acid sequences of putative protein domains, Gag, protease, reverse transcriptase, RNase H, and integrase, are aligned. Seven conserved subdomains of reverse transcriptase are underlined (Xiong and Eickbush 1990). The amino acid sequences of CsRn1 domains are compared with those of other LTR retrotransposons: Kabuki of Bombyx mori (AB032718), undescribed LTR retrotransposon of Drosophila melanogaster on AE003787, Sushi of Fugu rubripes (AF030881), and Osvaldo of Drosophila buzzatii (Z46728). For the CHCC Gag domain, only those of Kabuki and AE003787 are used for the alignment, since the other elements have a CCHC Gag motif instead of CHCC. Sequence identities with CsRn1 are highlighted in black, while similar residues are highlighted by gray boxes.

98.7%, which slightly differed in LTRs (98.3%) and in internal coding regions (98.9%) (table 1). In three copies (CsRn1-4, CsRn1-7, and CsRn1-52), the reading frames were retained, while in others they were corrupted by indels and/or stop codons introduced by base substitutions. All copies were bounded by TSDs of 4 bp and adjacent unique single-copy regions, supporting the heterogeneous genomewide distribution of CsRn1.

A sequence alignment of 14 full-length CsRn1 copies showed base substitutions shared by more than two copies at numerous positions (90 of 5,023 bp), which showed clear subdivisions within the CsRn1 elements. For example, two nucleotide pairs at positions 2666 and 2692 divided CsRn1 copies into two groups, designated group G, with diagnostic bases of G‘T, and group C, with C‘C. The two groups were further distinguished by nucleotide pairs observed at eight other positions, although the bases at these positions were less distinctive (data not shown). This, together with the finding of well-conserved sequences at the remaining positions, demonstrated that the members of group G formed a clonal lineage originating from a common progenitor copy. The three diagnostic bases at positions 66, 4240, and 4619 suggested further divergence of this group into two subgroups (fig. 6).

Short branch lengths relating the members of group G in a phylogenetic tree (average length = 0.001; fig. 6) (Medstrand and Mager 1998) and nearly identical LTR sequences of the copies (table 1) (Dangel et al. 1995) suggested that the CsRn1 copies of this group have recently been replicated. Together with the fact that the major fraction of the CsRn1 elements belong to group G (7 of 14 sequenced copies), these results demonstrate a role of the elements belonging to group G as active master copies for the multiplication of CsRn1 elements. The copies of group C also had numerous diagnostic bases, but they were heterogeneous and showed higher levels of sequence divergence when compared with group G (fig. 6). Thus, these divergent copies were thought to be inactive variant forms of CsRn1 that expanded prior to the expansion of group G.

The Mobile Activity of the CsRn1 Master Copy

The genomic distribution of the element was found to be heterogeneous among individuals of C. sinensis, which supports the recent expansion of CsRn1 (fig. 7A). Based on this finding, we attempted to examine the presence of CsRn1 transcripts in the total RNA molecules extracted from the adult worms by Northern blot analysis but failed to detect any signals when the blots were probed with P-B13 probe (data not shown). We then performed PCRs with higher sensitivity using a cDNA library as template and four primer sets covering the full-length of mRNA transcripts (see fig. 7B and its legend). As shown in figure 7B, 3′-end and P-B13 regions were well amplified, whereas 5′-end and RT regions were not amplified or weakly amplified, possibly due to the incomplete extension of cDNAs from 3′ ends during the construction of the cDNA library.
FIG. 5.—Intragenomic distribution and copy number of CsRn1. A, Genomic Southern blots showing that CsRn1 is highly reiterated throughout the Clonorchis sinensis genome. Restriction endonucleases used for the digestion of DNAs are indicated at the top. Arrows in lane 1 (AccI digests) indicate two prominent bands of 950 bp and 1,059 bp corresponding to the expected sizes of AccI digests spanning the LTRs and internal sequences at the 5' and 3' regions, respectively (see fig. 2). The positions of DNA size standards (in kb) are shown on the right. B, Dot-blot analysis for determining the copy number of CsRn1. Each of two membranes was dotted in duplicate with varying amounts (µg) of genomic DNAs as shown on the right. The probes used for hybridization are indicated at the top: P-B13 for CsRn1 (see fig. 2) and Cys-Pro for cysteine protease as a single-copy control. The signal intensities between the two blots were compared to estimate the copy number of CsRn1.

The PCR products of 3'-end regions were cloned and sequenced, and 12 randomly chosen clones showed six different sequences. The 3' ends of CsRn1 transcripts lay within 141–146 bp downstream of the 5' end of the 3' LTR and about 15 bp downstream of the presumed AATACA poly A signal sequence. When compared with genomic copies of CsRn1 elements, the transcripts shared the same bases with the copies of group G at diagnostic substitution positions (data not shown). A homology search of CsRn1 sequence against dbEST of GenBank revealed that a yet-unidentified CsRn1-like element is also expressed in Schistosoma mansoni, one of the well-studied trematodes (fig. 7C). Since the expansion of retrotransposons is largely restricted to germ cells and/or early embryonic stages, the frequency of CsRn1 amplification could not be determined by Northern blot analysis, in which the total RNA molecules from the whole organism were used. However, the polymorphic distribution of CsRn1 among individuals and the presence of its mRNA transcripts suggest constant, uninterrupted expansion of the element responsible for the high copy number.

The Evolutionary Relationship of CsRn1 with Other Retrotransposons

The amino acid sequence of RT encoded in CsRn1 showed strong homology with that of RT in many Ty3/ gypsy-like LTR retrotransposons, particularly in Kabuki (identity value of 63%) and AE003787 (58%), and a similar pattern of homology was also found with the sequences of IN (fig. 4). Moreover, the nucleotide sequence of CsRn1 exhibited homology with the DNA sequences of the two elements over a relatively long
FIG. 6.—Phylogenetic analysis of the full-length CsRn1 elements. The analysis was conducted using PHYLIP based on an alignment of full-length nucleotide sequences. The tree was constructed using the neighbor-joining algorithm and was unrooted. The number at a particular node indicates its percentage of appearances in 1,000 bootstrap replicates, and only values of >60% are indicated. Groupings supported by the bootstrap analysis, as well as by the diagnostic bases, are marked as thicker branches (see text).

In a previous report, Malik and Eickbush (1999) divided Ty3/gypsy-like LTR retrotransposons into eight distinct clades. As shown in figure 8, the members of the eight clades were well separated in a tree constructed by the UPGMA method. Interestingly, however, CsRn1 formed a previously undetected, tightly conserved clade with Kabuki and AE003787. A similar clustering pattern was observed in a tree constructed with the neighbor-joining algorithm (data not shown), and the statistical significance of the branching points was well supported by bootstrap analysis. As members of a new clade (designated the CsRn1 clade), the three elements shared a number of common features, such as a CHCC Gag mo-

FIG. 7.—Mobile activity of CsRn1. A, Southern blot analysis of CsRn1 with genomic DNAs separately extracted from individual worms (presented as numbers on the top) and P-LTR probe (see fig. 2). HindIII restriction endonuclease was used for the digestion of the DNAs. Several polymorphic bands among individuals introduced by recent expansion of CsRn1 can be seen between 0.56 and 1.4 kb. B, Amplification of 5’-end, reverse transcriptase (RT), P-B13 marker, and 3’-end regions of CsRn1 from a cDNA library of Clonorchis sinensis by PCR methods. Primers used in each PCR reaction are as follows: T3 and 5LTR-R (5’-CGACTAAATCGCTGAATC-3’) for the 5’-end region; RT-F (5’-GACGAAAGGTCCACCTGTC-3’) and RT-R (5’-CCCAATGACCTCCCAACGTC-3’) for RT; Pro-F (5’-CTCTGTTGAGCGTTCCATC-3’) and Pro-R (5’-CTGCTAGACCGTTGACCGTG-3’) for P-B13; and T7 and 3LTR-F (5’-GAAACTTGAAGTGAGCAAC-3’) for the 3’-end region. M = 100-bp size marker. C, Homology search of CsRn1 genomic sequence against the dbEST of GenBank databases. The full-length CsRn1 and three expressed sequence tags (ESTs) of Schistosoma mansoni with the highest match are presented. The hatched boxes indicate the matched regions, and their positions in the CsRn1 sequence are shown, together with the homology values. The ESTs of S. mansoni are presented with accession numbers in the databases and were obtained from cDNA libraries of male (AI977543) and female (AI975406 and AI976475) adult worms of the trematode.
Blastopia, D. melanogaster (Z27119); ponicum (Y12432); Cft1, Cladosporium fulvum (AF051915); sposons of Arabidopsis thaliana; AF026205 and U88169, undescribed retrotransposon of D. melanogaster; AF0262041, undescribed retrotransposon of Arabidopsis thaliana; AF026205 and U88169, undescribed retrotransposons of Caenorhabditis elegans; Athila, A. thaliana (AB005248); Blastopilia, D. melanogaster (Z27119); Cer1, C. elegans (U15406); Cft1, Cladosporium fulvum (AF051915); Deul, Ananas comusus (Y12432); Cyclops, Vicia faba (AB007466); Gulliver, Schistosoma japonicum (AF243513); Gypsy, D. melanogaster (M12927); Kabuki, Bombyx mori (AB032718); Mag, B. mori (S08405); MurY1, Tricholoma matsukake (AB028236); Mdg1, D. melanogaster (X59545); Mdg3, D. melanogaster (X59508); Osvaldo, D. buzzatti (A133521); Rire7, Orzya sativa (AB032325); Sashi, Fugu rubripes (AF030881); Ted, Trichoplusia ni (M32662); Tf1, Schizosaccharomyces pombe (M38526); Tom, D. ananassae (M32662); Ulysses, D. virilis (AF056940); Ty3-2, Saccharomyces cerevisiae (S35377); Ulysses, D. virilis (X56645); Woot, Tribolium castaneum (U09586); Yoyo, Ceratitis capitata (U60529); ZAM, D. melanogaster (AJ000387).

During an evolutionary time, a particular subset of retrotransposons expands differentially, rather than simultaneously, from other variant subsets with selective advantage for expansion (Clough et al. 1996). Thus, a recently expanded master copy can be distinguished as the subset with the largest population (Boissinot, Chevre, and Furano 2000) and with low levels of sequence divergence among its members (Medstrand and Mager 1998). In addition to these characteristics, high sequence identity between flanking LTRs of individual copies can be accepted as a hallmark of their recent integration in cases of LTR retrotransposons, since a pair of LTRs use the same sequences as templates for their replications (Dangel et al. 1995). CsRn1 copies of group G satisfied all of these criteria (table 1 and fig. 6), suggesting the role of group G as an active master copy, and the preserved mobile activity was confirmed by the uncorrupted ORF, heterogeneous distribution among individual genomes (fig. 7A), and the presence of mRNA transcripts of CsRn1 (fig. 7B).

The coding capacity of CsRn1 suggests that the element is a member of metaviruses, which have no env gene (Pringle 1999). Although no significant homology to other LTR retrotransposons at the nucleotide sequence level was found throughout the whole unit, sequence motifs for the synthesis of double-stranded cDNA (PBS and PPT) showed strong identity with those of Kabuki (B. mori) and AE003787 (D. melanogaster) (fig. 3). Moreover, the amino acid sequences in the unusual CHCC Gag motif, RT, and IN were well conserved in the three elements (fig. 4). With these shared properties, a phylogenetic analysis suggested that CsRn1 formed an ancient, previously undetected clade of Ty3/gypsy-like LTR retrotransposons found in insects and trematodes (CsRn1 clade; fig. 8). The nucleotide sequences of Kabuki and AE003787 in their putative ORF regions are corrupted, which inactivates the elements and gives rise to low copy numbers (the number of Kabuki was estimated as eight in the haploid genome; Abe et al. 2000). However, several copies of CsRn1 were uncorrupted and

Discussion

In the present study, we characterized CsRn1 from the genetically polymorphic regions among individual C. sinensis worms as the first member of uncorrupted LTR retrotransposons found in the phylum Platyhelminthes. The full-length CsRn1 encodes a single uninterrupted ORF which resembles pol protein of Ty3/gypsy-like elements. A phylogenetic analysis showed that CsRn1 is a member of a distinct, previously undetected clade of LTR retrotransposons which exhibit definitive characteristics such as highly conserved PBS (tRNA^p^) and PPT, a highly conserved and unusual CHCC Gag motif, a TSD of 4 bases, and strong similarity of functional protein domains.

For expansion, retrotransposons are transcribed by host RNA polymerase and then reverse-transcribed by their own reverse transcriptase. Because these two enzymes have no proofreading capacity (Varmus and Brown 1989), cDNAs produced during the process of transposition tend to acquire random base substitutions. These base substitutions frequently inactivate the progeny copies, and the resulting “dead-on-arrival” (DOA) copies with no mobile activities are likely to be subjected to neutral evolution, through which the accumulation of sequence variations within the DOA copies is accelerated (Petrov, Lozovskaya, and Hartl 1996). Thus, together with the low fidelity of RT, the neutral evolution of the inactive DOA copies may have an additional effect on sequence divergence among individual genomes of a species. In this study, a variable genomic region among individuals was successfully identified by AP-PCR-based RAPD analysis as CsRn1 LTR retrotransposon-related sequence.
showed maintained mobile capacity (fig. 7), suggesting that the element might have acquired high copy numbers in *C. sinensis* genomes (more than 100 per haploid genome; fig. 5) through its continuous expansion.

Although they had a phylogenetically close relationship, differences in the numbers of ORFs and no significant homology in nucleotide sequences were observed among the members of the new clade, which reduces the probability of horizontal transfer between insects and trematodes in the recent past. Thus, it is likely that these elements evolved from a common ancestor that was present in the common progenitor of insects and trematodes or was transferred from insects to trematodes, or vice versa, during the early stage of the divergence of the two taxa. However, the possibility of horizontal transfer between insects and trematodes is uncertain because few cases, only within similar species, have been reported (Gonzalez and Lessios 1999; Jordan, Matynina, and McDonald 1999). The isolation of further elements belonging to the *CsRn1* clade in insects and trematodes or the finding of an errantivirus(s) that is phylogenetically related to the clade (Malik and Eckbush 1999) will be helpful in understanding the detailed evolutionary course among the members of the clade.

Only a few data on LTR retrotransposons are available for the Platyhelminthes, despite the large content of repetitive elements in their genomes (see Regev, Lamb, and Jablonka 1998 and references therein). Model organisms such as *D. melanogaster*, *A. thaliana*, and *S. cerevisiae*, commonly used in previous studies, have genome structures with relatively low complexity and repetitive elements with low copy numbers, which makes it difficult to estimate the actual significance of retrotransposons in the complex genome of the animal. Thus, the present study using a trematode provides advantages for the study of retrotransposons with small but complex genomes. We confirmed the presence of partial sequences similar to those of *CsRn1* in other trematodes, *S. mansoni* (fig. 7C) and *Paragonimus westermani* (unpublished data). These results reflect the presence of a unique LTR retrotransposon family belonging to the *CsRn1* clade in the lower animal taxa which may play significant roles in the evolution of genomes. Our results concerning an active LTR retrotransposon and its related elements will broaden the current knowledge on LTR retrotransposons and provide a clue for further studies on the evolutionary origin of diverse reverse-transcribing elements.

**Supplementary Material**

All of the nucleotide sequences described in this article were deposited in GenBank in linked form to each nucleotide set, of which accession numbers are presented above, and the alignment of *pol* proteins used for phylogenetic analysis in this work was deposited in linked form to the nucleotide sequence of *CsRn1*-4 (AY013569).

A sequence alignment of multiple *CsRn1* copies was deposited in GenBank in linked form to each nucleotide set, of which accession numbers are presented above, and the alignment of *pol* proteins used for phylogenetic analysis in this work was deposited in linked form to the nucleotide sequence of *CsRn1*-4 (AY013569).

**Acknowledgments**

This work was supported by a grant from the Korea Science and Engineering Foundation (KOSEF) (number 1999-2-208-004-5).

**LITERATURE CITED**


Kenneth Wolfe, reviewing editor

Accepted April 4, 2001