Skipper, an LTR retrotransposon of Dictyostelium

Ping Leng*, David H. Klatte, Gerald Schumann¹, Jef D. Boeke¹,* and Theodore L. Steck

Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637, USA and
¹Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Received November 6, 1997; Revised and Accepted March 4, 1998

ABSTRACT

The complete sequence of a retrotransposon from Dictyostelium discoideum, named skipper, was obtained from cDNA and genomic clones. The sequence of a nearly full-length skipper cDNA was similar to that of three other partially sequenced cDNAs. The corresponding retrotransposon is represented in ~15–20 copies and is abundantly transcribed. Skipper contains three open reading frames (ORFs) with an unusual sequence organization, aspects of which resemble certain mammalian retroviruses. ORFs 1 and 3 correspond to gag and pol genes; the second ORF, pro, corresponding to protease, was separated from gag by a single stop codon followed shortly thereafter by a potential pseudoknot. ORF3 (pol) was separated from pro by a +1 frameshift. ORFs 2 and 3 overlapped by 32 bp. The computed amino acid sequences of the skipper ORFs contain regions resembling retrotransposon polyprotein domains, including a nucleic acid binding protein, aspartyl protease, reverse transcriptase and integrase. Skipper is the first example of a retrotransposon with a separate pro gene. Skipper is also novel in that it appears to use stop codon suppression rather than frameshifting to modulate pro expression. Finally, skipper and its components may provide useful tools for the genetic characterization of Dictyostelium.

INTRODUCTION

Retroelements are DNA sequences ubiquitous in eukaryotes from yeast to humans as either indigenous retrotransposons or infectious retrovirus genetic elements. They replicate through an RNA intermediate and integrate into the chromosomal DNA of the host cell using enzymes which they encode, including a reverse transcriptase and, in some cases, capsid proteins, a protease and an integrase. The latter proteins are encoded only by the retrotransposons that have long terminal repeats (LTRs) and hence more closely resemble retroviruses. The gene products are initially translated as polyproteins, often involving one (or rarely two) translatable frameshift(s). The polyproteins are then assembled into a precursor virus-like particle, in which the primary translation products are processed into functional products by the encoded protease. Multiple copies of retrotransposons, or retroviral proviruses, may be present in the host genome; some of these may be functional (able to carry out all replicative activities) and others defective in some way. There are many different types of retrotransposons, and their biology has been extensively reviewed (1).

Four transposon families have been previously described in the social amoeba, Dictyostelium discoideum, Tdd-2 and 3 (2,3), DIRS-1, also called Tdd-1 (2,3), Dirs-1, also called Tdd-1 (4) and DRE (5,6). Tdd-2 and 3 elements have not been well-characterized but appear to lack LTRs; they do have A-rich termini (suggesting they might be polyA retrotransposons, also referred to as non-LTR retrotransposons), but little else is known of their structure. DRE elements belong to the polyA class of retrotransposons based on alignment of their RT sequences with those of other retroelements; DIRS-1 has an unusual terminal structure, with what appear to be inverted LTRs, and its RT sequence is atypical and has not been clearly classified. In the present study, we describe a new Dictyostelium retrotransposon, skipper, which belongs to the class of LTR elements exemplified by Ty3 in yeast and gypsy in Drosophila. Its overall structure is shown in Figure 1.

MATERIALS AND METHODS

Cell culture, bacterial strains and vectors

Dictyostelium discoideum strains AX-2 and AX-3 were grown as amoebae on axenic medium as previously described (6). The Lambda ZAP II host strain, XL1-Blue MRF was employed to prepare the cDNA library (Stratagene, LaJolla, CA). SOLR™ was used in the ExAssist™/SOLR™ system for excision of the phagemid (Stratagene, LaJolla, CA). DH5α was used as the host strain for all other procedures.

*To whom correspondence should be addressed. Tel: +1 410 955 2481; Fax: +1 410 614 2987; Email: jef_boeke@qmail.bs.jhu.edu

+Present address: Department of Molecular and Cell Biology, Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen AB25 2ZD, UK

§AF017040–AF017047, AF049229 and AF049230

The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors.
Total RNA was extracted from axenic amoebae, grown to stationary phase (1.5–2 × 10^7 cells/ml) and purified using an Ultraspec™-II RNA kit (Biotex, Houston, TX). RNA samples (20 mg) were analyzed by electrophoresis on 0.8% agarose gels containing 2.2 M formaldehyde and transferred to nitrocellulose membrane (Schleicher & Schuell, NH). RNA blot hybridization was performed by standard procedures (8) using random primed 32P-radiolabeled cDNA as a probe. Hybridization was carried out in 6× SSPE containing 50% formamide, 10% dextran sulfate, 5x Denhardt’s reagent, 0.5% sodium dodecylsulfate (SDS) and 100 mg/ml denatured fragmented salmon sperm DNA. Wash conditions were as follows: two washes in 2× SSPE + 0.2% SDS at 50°C for 15 min, then two more washes in 0.1× SSPE at 60°C for 15 and 5 min.

Genomic DNA blot analysis

Aliquots (3.6 µg) of genomic AX-2 and AX-3 DNA (kindly provided by P.Devreotes) were digested with the restriction endonucleases SpeI and SpeI/NcoI. Genomic DNA fragments were fractionated on a 0.9% agarose gel and transferred to a GeneScreen Plus (DuPont) membrane. All DNA probes were labeled for hybridization with [γ-32P]dCTP by random priming (9).

Construction of cDNA library

Total RNA was prepared as for RNA blotting. Poly(A)+ RNA was isolated using an mRNA purification kit (Pharmacia Biotech, Uppsala, Sweden). cDNA synthesis was described as in ref. 8. The resulting DNA was size selected on a 0.8% low melting point agarose gel and blunt ligated to EcoRI adapters (Promega, Madison, WI). The corresponding cDNA library was established in the Lambda ZAP II vector (Stratagene, La Jolla, CA) according to the manufacturer’s directions.

Sequencing strategy

Digestion with exonuclease III and mung bean nuclease was employed to generate a set of nested subclones (10). Double-stranded DNA was sequenced in one of two automated DNA sequencing facilities: the Howard Hughes Medical Institute or the University of Chicago Cancer Research Center DNA Sequencing Facility. In the first case, sequencing reactions were performed using the PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit (Perkin Elmer Corporation, Foster City, CA) according to the manufacturer’s directions. In the second case, reactions were performed by the facility. Gaps were filled by sequencing from the undigested parent clone using flanking custom primers (Integrated DNA Technologies, Inc., Coralville, IA).

Identification of LTR-sequences

PCR on genomic AX-3 DNA. Various primer pairs reading outward from the cDNA sequence termini were combined in various ways to provide evidence for the presence of full-length genomic elements with LTRs (see Fig. 2 for strategy). Four independent PCR reactions on genomic DNA from *D.discoideum* strain AX-3 were performed using the primer combinations JB1334/ JB1331 and JB1334/JB1330. The resulting 600 bp PCR-products were subcloned into plasmid pCRII™ (Invitrogen) giving rise to the plasmids pGS.405.1 and pGS.409 (primers JB1334/JB1330; GenBank AF017040) and pGS.414 and pGS.413 (primers JB1334/JB1331).

**Figure 1.** Overall organization of the *skipper* element. Gag corresponds to retrotransposon and retroviral capsid/nucleocapsid domains. Pro encodes the protease (PR) and Pol encodes the RT and IN domains. As with other *gyrps*/Ty3 family retrotransposons, the overall organization is LTR-GAG-PR-RT-IN-LTR. A single stop codon separates gag and pro, while a +1 frameshift, with a short reading frame overlap of 11 amino codons, separates pro from pol. Black arrows indicate LTRs which occupy positions 1–390 and 6609–6998. Gag corresponds to nucleotides 1132–2280. Pro is from 2281–3249. Pol is from 3215–5665. The cDNA fragment included in prD41 begins at position 331 and ends at position 6445; its GenBank accession number is AF017040. The accession number of the composite sequence diagrammed here is AF049230.

**Figure 2.** Identification of the complete structure of the *skipper* element. Based on the sequence analysis of the *skipper*-specific cDNAs (prD33, prD41, prD44, prD72), we designed oligonucleotides JB1331 and JB1334, which are specific for the 3’- and 5’-ends of the isolated cDNA from prD41. PCR performed on genomic DNA from *D.discoideum* AX-3 resulted in a 600 bp product. The synthesis of such a PCR product with this combination of primers suggested the presence of LTRs. The PCR product was cloned and used to screen a genomic HindIII-plasmid library from *D.discoideum* AX-2. The isolated *skipper*-LTR containing clones carried either the 5’ or the 3’ LTR. LTR sequences of seven different clones were analysed and the consensus was merged with the cDNA sequence (GenBank AF017040). The structure of the complete *skipper* element is also indicated by dashed lines and a grey LTR.
The oligonucleotide primers were as follows (the coordinates listed below and throughout this paper are nucleotide positions from the composite intact skipper sequence; GenBank AF049230).

*JB1330 5′-CAATATCTCCAAAGCAAAACATCC (pos. 6402–6427);
JB1331 5′-CAAGACTTCAATCAGGAAAGAAC (pos. 6419–6445);
JB1334 5′-CAATGCTTGAATGAAACAGAG (pos. 6491–6517);
JB1605 5′-CTAAGATCAATGATCATACTATCAG (pos. 6510–6538);
JB1645 5′-CTGGTATCCCTAATCTGAGGATCG (pos. 517–519);
JB1646 5′-CAAGTGTAACGAGTAAATCTTACAG (pos. 6483–6510)*

Screening of genomic AX-2 library. Recombinant clones carrying either the 5′- or 3′-ends of different skipper elements were isolated by screening a genomic library from *D.discoideum* strain AX-2. The plasmid library was constructed by ligating HindIII-restricted genomic DNA into pGEM-7Zf(–) plasmids (Promega), kindly provided by T.Dingermann (Johann-Wolfgang Goethe Universität, Frankfurt/Main, Germany). Recombinant plasmids were transformed into the *Escherichia coli* DH5α strain and screened for skipper LTRs using the labeled 600 bp EcoRI fragment from pGS.405.I as a probe.

Sequence analysis of identified clones was performed using the oligonucleotides JB1645 (reads towards the 5′-end of LTR), JB1604 and 1646 (read towards the 3′-end of LTR). In order to confirm the junction between the 3′-end of the skipper cDNA and the 5′-end of the LTR-containing PCR product obtained from genomic DNA, the genomic clones carrying 3′-ends of skipper were sequenced with the oligonucleotide JB1605.

**Computer analysis**

Sequence analysis was performed using the X-windowed UNIX form (version 8.1-UNIX) of the Wisconsin Package (‘GCG’) (Wisconsin Package, 1994). We did evolutionary comparisons using the core domain of RT; we performed a multiple sequence alignment with several known gypsy/Ty3 group retrotransposons and HIV for comparison (not shown) and calculated an evolutionary tree using the unweighted-pair-group method (UPGMA) (11). For both the alignment and the tree, we observed results that were essentially the same as a subset of previously published data (7), but with the addition of skipper. Skipper clearly fits into the Ty3/gypsy family.

**RESULTS**

cDNA clones of a novel retrotransposon

Four clones later shown by DNA sequencing to contain the skipper element were accidentally selected from a cDNA library prepared as described in Materials and Methods. The probe sequence, not present in skipper, must contain fortuitous complementarity to skipper RNA.

The DNA sequences of overlapping exonuclease digest fragments of the skipper cDNA clone 41 were analyzed in both directions. Overlaps were determined and the sequence gaps were filled by performing sequencing reactions with custom primers. A partial (~80–90%) sequence of a second clone, skipper clone 33, was generated in one direction only; gaps were not filled. It was nearly identical to the first and is not shown. Two other clones (44 and 72) have identical 3′-ends to skipper-41; however, relative to skipper-41, they lack 454 and 548 bp of sequence, respectively, from their 5′-ends. While clones 44 and 72 were not analyzed in their entirety, their 5′- and 3′-ends matched those of skipper-41 (not shown). Also, all of the clones were analyzed in the regions corresponding to the ORF boundaries and these regions were conserved.

The skipper cDNA clone 41 insert was 6115 bp in length (GenBank AF017040). Typical of *Dictyostelium*, its sequence was AT rich: 67.7% A+T and 32.3% G+C. Paradoxically, all four cDNA clones lacked poly-A tails, even though they came from an oligo-dT primed library, and shared an identical 3′-end sequence.

**Characteristics of the 5′ sequence of skipper-41 cDNA**

The initial 801 bp of the cDNA contained nine ATGs in closed reading frames; i.e., followed within 50 bp by an in-frame stop codon. Eukaryotic translational mechanisms are not fully understood, but the widely accepted scanning model (12) holds that translation is initiated by the first (or sometimes the second) AUG, generally within 200 nt of the 5′-end. However, there are exceptions to this pattern (e.g., picornaviruses), and well documented cases of internal initiation of translation exist (13). Clearly, such an mRNA could not be efficiently translated via the conventional scanning mechanism.

A 53 bp sequence near the 5′-end of the cDNA (position 625–677) is 77% identical to a portion of the Ty3 promoter region from yeast and includes a conserved TATA consensus sequence (14). Following this promoter-like region was a series of 43 bp repeats (position 691–965). The first three were consecutive and identical (position 691–819). They were followed by a fourth repeat of 95% identity (position 820–862). This was followed by 60 bp of unrelated sequence, followed by a sequence which was 93% identical to the first three repeats (position 922–965). Following the repeat region was a 156 bp sequence which was 97.4% A+T.

**LTR identification and characterization**

The cDNA sequence/structure was not entirely consistent with the identification of skipper as an LTR retrotransposon because the inferred RNA sequence lacked the terminal repetitive characteristic of LTR retrotransposons. However, the cDNAs might have been incomplete. We therefore reasoned that if LTRs were present in the host genome, we could clone them based on the terminal sequences of the cDNA. We tested for the presence of LTRs in genomic *D.discoideum* DNA using a PCR strategy (see Materials and Methods and Fig. 2 for strategy). A sense primer corresponding to the 3′-end of skipper cDNA and an antisense primer corresponding to the 5′-end of skipper cDNA were synthesized and used in a PCR reaction with genomic DNA as template. A major product of ~600 bp was obtained. The products of such PCR reactions were cloned and clones from four independent PCR reactions were isolated, sequenced and aligned. The alignment showed that all of the variation was limited to what was later determined to be the LTR segment of these fragments. The fact that the PCR products from a single primer pair were not completely homogeneous, implies the existence of multiple, slightly different elements in the genome. However, it may be that some of the observed sequence variations in this experiment arose as artifacts of the PCR. The existence of minor variants of the LTR is a common feature of retrotransposon families (15). At this point we had evidence for an LTR-type structure for skipper, but...
it was not possible to predict with certainty the terminal sequences of the element.

To determine the precise 5′ and 3′ boundaries of the LTRs, we cloned genomic DNA fragments containing skipper LTRs and sequenced the termini; this allowed unambiguous definition of the ends of the LTRs (Fig. 3). All of the clones corresponded to sequenced the termini; this allowed unambiguous definition of cloned genomic DNA fragments containing the element.

The sequences flanking the six skipper termini were compared to the Genbank database using BLAST. Only two of the six had a significant match, which were to the DIRS-1 retrotransposon. DIRS-1 is known to transpose into itself (16); this result suggests that DIRS-1 can also be a target for other retroelements. Skipper inserted in the orientation transcriptionally opposite that of DIRS-1 in both cases; both insertions were in the region of overlap between ORF2 and ORF3 of DIRS-1.

Based on the experimentally determined LTR termini, the positions of the primer binding site (PBS) at which minus strand reverse transcription initiates, and polypurine tract (PPT) at which plus strand reverse transcription initiates, could be predicted. The putative PBS of retrotransposons is usually complementary to a specific host tRNA. However, we have not been able to identify a known D.discoideum tRNA that is fully complementary to the skipper PBS sequence through database searches. In fact, the skipper PBS sequence, CTTTTTTTTTCT-AATT (position 391–408), is extremely unusual, beginning with a run of 12 consecutive pyrimidines, mostly Ts. As no known Dictyostelium tRNA matches this sequence (and complementarity to the conserved CCA 3′ terminus of tRNAs is lacking from the skipper PBS), it is likely that priming of minus strand reverse transcription of this element occurs by some other means. On the other hand, the skipper PPT sequence is quite typical, consisting of 18 consecutive purines immediately preceding the 3′ LTR. A composite DNA sequence (AF049230) incorporates the consensus LTR sequence obtained from sequencing the four independent LTR PCR products, the seven genomic LTR sequences and the cDNA clone 41 sequence.

**Analysis of hypothetical protein regions**

Codon usage frequencies in ORF1 and ORF2 were compared with a D.discoideum codon preference table that averages 60 611 codons from known Dictyostelium ORFs (17) using the GCG Codon Preference program (not shown). Relatively good conservation of Dictyostelium codon usage was evident for most of the ORF sequences.

Database searches with GCG Blast identified regions of skipper which resembled retroelement coding regions and, in particular, LTR retrotransposons (not shown). The inferred amino acid sequences of the three ORFs of skipper were then compared with those of known retroelement protein domains.

**Nucleic acid binding domain.** The gag gene of retroviruses is not generally well conserved, but it contains a small putative nucleic acid binding domain in the nucleocapsid (18–21). This region codes for a characteristic amino acid sequence: C-X2-C-X4-H-X4-C. Typically, this sequence occurs twice in retroviruses and once in retrotransposons such as Ty3 (14). Skipper contains a single sequence of this type (Fig. 4a).

**Protease.** The polyproteins encoded by retroelements are processed by a protease to produce the individual mature proteins (22,23). A core highly conserved peptide sequence (hydrophobic-hydrophobic-H-D) is found in ORF2 of both in numerous retroelement proteases, and in other aspartyl proteases is also found in ORF2 of skipper (Fig. 4b).

**Reverse transcriptase.** Reverse transcription of retroelements utilizes polymerase and ribonuclease H activities. Sequences specific for these activities occur characteristically in retroelement reverse transcriptase coding regions (24,25). An alignment of this region of skipper ORF3 with the corresponding region from Ty3 (14) indicates that these sequences are 38% identical at the amino acid level (Fig. 4c). Skipper contains several conserved amino acids characteristic of the core domain of retroelement reverse transcriptases (26).

**Integrase.** An integrase is found downstream of the reverse transcriptase in the gypsy/Ty3 class of retrotransposons; such an integrase-like domain is found in skipper. Like all other integrase sequences, the N-terminus contains a consensus zinc-finger-like HH-CC motif. The predicted amino acid sequence for the core integrase (so called D-D35E) region of skipper is 38% identical to the Ty3 integrase sequence (Fig. 4d). In addition, there are
Figure 4. Peptide alignments and relationship tree. (a–d) Peptide alignments. The symbols between aligned sequences indicate their relationship; a bar (I) indicates an identical residue, a colon (:) indicates a conserved amino acid, and a period (.) indicates a semi-conserved residue. In the sequences, gaps necessary for the alignment are indicated by periods (.). (a) Alignment of the core nucleic acid binding domain in skipper with the same region from Ty3 (corresponding to position 2173–2226). (b) Conserved region of the protease domain in skipper (corresponding to position 2881–2928) aligned with the corresponding region in Ty3. The most highly conserved aspartyl protease domain motif (hydrophobic-hydrophobic-D-T/S-G-A/S) is underlined. (c) The catalytic domain of the reverse transcriptase of skipper (corresponding to position 3596–4129) aligned with equivalent section of Ty3. These sequences are 33% identical. Retroelement reverse transcriptase motifs are in bold face and underlined. (d) The integrase core domain of skipper (corresponding to position 5042–5626) aligned with the Ty3 integrase sequence, with which it is 38% identical at the amino acid level. Retroelement integrase motifs are in bold face and underlined. (e) Evolutionary tree, based on UPGMA alignments of the reverse transcriptase core domain in (c), showing the relative relationships of members of the gypsy/Ty3 family of retrotransposons, with Ty1 for comparison.

several amino acid sequence motifs which are conserved in the integrases and transposases of retroviruses, retrotransposons and bacterial insertion sequences (27).

Evolutionary analysis. Using the core domain of RT, we performed a multiple sequence alignment with several known gypsy/Ty3 group retrotransposons and HIV for comparison (not shown) and calculated an evolutionary tree using the unweighted-pair-group method (UPGMA) (11). Our results for both the alignment and the tree were essentially the same as a subset of previously published data (28), but with the addition of skipper. Skipper clearly belongs to the Ty3/gypsy family (Fig. 4e).

Arrangement of the open reading frames

The arrangement of ORFs in the skipper element is unique among the LTR retroelements. The first two ORFs, gag and pol, are separated by a single UGA stop codon. Pro and pol are separated by a +1 frameshift (Fig. 1). This arrangement of ORFs is maintained in all four isolates sequenced in these regions, and so is presumed to represent the normal structure of the skipper element rather than a mutant variant. The length of pro corresponds reasonably well to the size of some retrotransposon proteases and is nearly exactly the same length as the X-PR ORF of the following mammalian retroviruses: mouse mammary tumor virus (MMTV; GenBank M15122), human T-cell leukemia virus (GenBank L36905) and Mason-Pfizer monkey virus (GenBank M12349). All of these viruses encode PR as part of a separate ORF. In these viruses, however, the X-PR ORF is separated from each of its neighbors by –1 frameshifts (29). Gag and pol are separated by a single UAG stop codon in Moloney murine leukemia virus (Mo-MLV), and this stop codon is read through at an efficiency of 5%, resulting in the synthesis of Gag-Pol readthrough protein (30). It has been shown that to effect this readthrough, the identity of the stop codon is not
important, but a certain RNA secondary structure must be present immediately 3′ to the stop codon (31,32). This structure has the sequence characteristics of a pseudoknot, an RNA structure that is also downstream of a number of –1 Gag-Pol frameshift sequences in retroviruses (33–36), and in overlapping genes of other RNA viruses (37,38). We have found a sequence downstream of the skipper Gag-PR stop codon that can be folded into a pseudoknot structure and is likely to effect a readthrough process similar to that which occurs in Mo-MLV (Fig. 5).

The +1 frameshift separating pro and pol does not resemble the known +1 frameshifting sequences found in yeast Ty elements (39,40). However, we do note that, as is the case with both the Ty1 and Ty3 sequences, a rarely used codon is found in the upstream (PR) frame, just prior to its stop codon. The identity of this codon, UCCSer, differs from the rare codons that figure prominently in Ty1 and Ty3 +1 frameshifting.

DNA blot analysis

Retrotransposons are typically multicopy sequences; however, some exceptions exist, such as the Ty3 and Ty4 sequences, which are present in just one to four copies in some strains of Saccharomyces cerevisiae. We therefore carried out genomic DNA blot analysis of D.discoideum DNA digested with single-cut restriction enzymes (Fig. 6b). The blot was then hybridized with a skipper-specific probe that would recognize a single junction fragment from each transposon copy. This assumes that all elements will have the same restriction map as our consensus sequence, which probably is not true for every single skipper copy. Indeed, the pattern of bands obtained, mostly consisting of well-separated single-copy bands of >2.5 kb, suggests that the majority of the copies have similar maps. Such restriction site conservation is supported by an NcoI digest, which shows

![DNA Blot Analysis](image)

**Figure 6.** Abundance and expression of skipper elements in D.discoideum. (a) Restriction map of the skipper element. (b) Genomic AX-2 DNA was completely digested with SpeI and SpeI–NcoI, size fractionated and probed with the skipper-specific 1278 bp SalI–NcoI fragment (probe 1) from prdD41. BstEII digested phage λ DNA served as size marker (M). (c) Genomic AX-3 DNA was digested with NcoI, size fractionated and probed with the 2.56 kb BstXI fragment from prdD41 (probe 2) corresponding to the skipper 3′-end. Note co-migration with the internal fragment produced by prdD41. The other fragments are junction fragments. (d) RNA blot analysis of total RNA from AX-3 cells. The filter was hybridized with the plasmid prdD33, carrying skipper cDNA sequences. A predominant RNA in the range between 6 and 7 kb as well as two subgenomic RNAs are detected.
expression of a conserved internal restriction fragment of 3.67 kb (Fig. 6c). In digests using single-cut restriction enzymes such as *Sph*I, ~15–20 fragments of different sizes hybridized. Similar results were obtained with a *Sph*I–*Nco*I double digest, which resolved the fragments a little better. We estimate that *skipper* is present in ~15–20 copies per genome.

**Expression of the *skipper* element**

Are *skipper* elements transpositionally active? A first step in answering this question is to investigate whether they are expressed. The fact that multiple cDNA clones derived from *skipper* were accidentally isolated during a screening project suggests the existence of a rather abundant RNA. Thus, we examined total RNA preparations of *D. discoideum* by RNA blot analysis (Fig. 6d). A prominent band estimated to be ~6 kb in length was observed and the strength of the hybridization signal was such that the band was prominent in a 2 h exposure, consistent with transcript abundance. Two minor bands (probably corresponding to degradation products but possibly to discrete subgenomic RNA species) were also observed. The electrophoretic mobility and the uniformity of the hybridization signal suggest that this band represents an LTR-to-LTR transcript that is a good candidate for both the mRNA and the reverse transcription substrate for retrotransposition.

**DISCUSSION**

LTR retrotransposons have now been isolated from virtually all types of eukaryotes, so it is not surprising that one should turn up in *Dictyostelium*. *Skipper* contains all of the coding sequences associated with LTR retrotransposons, including a nucleocapsid (NC) motif, as well as robust predicted protein sequence homologies to retrovirus and retrotransposon PR, RT and IN proteins. Based on our analysis of the RT protein coding sequence, which is the most conserved coding sequence in the retroelements (25), the *skipper* element is a member of the Ty3/gypsy family of LTR elements, and its closest relative was the *Ulysses* element of *Drosophila melanogaster* (41).

*Skipper* represents the first retrotransposon we know of with a gag-pro-pol ORF organization. Its unusual genomic organization suggests that novel methods of modulating gene expression operate in this element. *Skipper* is only the third retroelement we know of in which a stop codon is found at a reading frame junction where a down-modulation of expression is expected. Like the analogous Mo-MLV case, the sequence predicts a pseudoknot structure downstream of the stop codon; whether this structure indeed forms and affects translational readthrough needs to be confirmed by direct experimentation. The protein that would result from reading through this stop codon would consist of Gag-PR, MMTV makes Gag-PR products by a single −1 frameshift event (29). Gag-PR proteins are also synthesized directly in the avian Rous sarcoma retrovirus (42) and from the major spliced 2.2 kb mRNA derived from the *copia* retrotransposon in *Drosophila* (43). Very recently, the sequence of *pCAL*, a new *Ty1/copia* family retrotransposon from *Candida albicans* was shown to contain a stop codon separating gag and pol (44). A +1 frameshift is found between the pro and pol (RT-IN) reading frames in *skipper*. Such +1 frameshifts are rather rare in retroelements. While it is likely that Pol protein is made by a +1 frameshift event from PR, the case cannot be made too strongly without direct experimental data: the overlap of the two ORFs might be fortuitous rather than representing a programmed reading frame shift. For example, even though the gag and pol gene equivalents in Hepatitis B viruses overlap, pol is separately translated from its own AUG (45,46). The spumaretrovirus, or foamy virus family, was recently shown to express its gag-pol mRNA by splicing and not by frameshifting (47). The human L1 retrotransposon lacks LTRs and has an intergenic region separating ORF1 and pol; this supports a reinitiation model (48). We cannot rule out independent translation of *skipper* pol or pro, for that matter, from minor spliced mRNAs or by reinitiation from scanning ribosomes. The former scenario seems unlikely though, given the uniformity of the *skipper* transcript, and low abundance of specific subgenomic transcripts. If pol is in fact made by programmed +1 frameshifting, as is the case in the yeast Ty elements, then it must be by a somewhat different set of factors than those used by Ty elements, as the appropriate codons for mediating Ty1-like and Ty3-like frameshifting are absent in the pro-pol overlap region.

Minus strand priming is also likely to be unusual in *skipper*. The peculiar primer binding site sequence is inconsistent with the conventional tRNA priming mechanism. Some elements, such as *copia* and Ty5 (49,50), use 5′ tRNA fragments as primers for reverse transcription; accordingly, the PBSs of these elements lack CCA complementarity. But in spite of the fact that many *Dictyostelium* tRNA sequences have been identified, we find no candidate primers in the database. TF1 and related elements also lack CCA complementarity but use yet another priming mechanism, involving base pairing of the PBS to the 5′-end of the transcript in a complex ‘pretzel’ structure followed by RNase H cleavage of the RNA to generate a primer 3′-end (51–53). The appropriate sequences for an analogous self-priming mechanism are missing from the *skipper* element. Thus, it is difficult to predict the likely minus strand priming mechanism for *skipper* with confidence. However, the T-rich sequence of the PBS raises one intriguing possibility: perhaps the polyA tail of *skipper* RNA might self-prime reverse transcription. Such a priming mechanism would raise special problems for primer removal, but experiments with the TF1 system suggest that its RNase H may in special situations have double-stranded RNase activity (RNase H*) which could remove the primer. Significant experimentation will be required to test this speculative model.

The *Dictyostelium* retroelement DRE is of great interest because of its unusual position-specific integration upstream of tRNA genes. None of the isolated genomic sequences flanking *skipper* resemble tRNA genes, and most of these flanking sequences were anonymous and novel. However, two of six sequences flanking the *skipper* insertions that we analyzed represent the sequences of yet another retroelement, Dirs1. Although Dirs-1 sequences comprise only 0.4% of the *Dictyostelium* genome, they represent one third of our small sample of insertion sequences, suggesting that Dirs-1 elements might represent a preferred target for *skipper* elements. Recent studies of yeast transposons indicate that many of them identify ‘safe havens’ in the genome that do not contain host genes, which allow them to increase their copy numbers without harming their host; preferred targets include tRNA upstream regions and silenced regions (54–57). Insertion into another retroelement sequence may represent an alternative form of safe haven exploited by the *skipper* element. It will be interesting to determine whether the other *skipper* targets might represent novel mobile elements.
### Table 1.

<table>
<thead>
<tr>
<th>Element</th>
<th>Type</th>
<th>Copy number</th>
<th>Element length (kb)</th>
<th>% of host genome</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ttd2</td>
<td>polyA?</td>
<td>20–30</td>
<td>4.7</td>
<td>0.2–0.3</td>
<td>(2)</td>
</tr>
<tr>
<td>Ttd3</td>
<td>polyA?</td>
<td>20–30</td>
<td>4.7</td>
<td>0.2–0.3</td>
<td>(2)</td>
</tr>
<tr>
<td>DIRS-1 (Tdd1)</td>
<td>atypical LTRa</td>
<td>40b</td>
<td>4.7</td>
<td>0.4</td>
<td>(4)</td>
</tr>
<tr>
<td>DRE-a</td>
<td>polyA</td>
<td>80–100</td>
<td>5.7</td>
<td>1.1</td>
<td>(5)</td>
</tr>
<tr>
<td>DRE-b</td>
<td></td>
<td>80–100</td>
<td>2.4</td>
<td>0.5</td>
<td>(6)</td>
</tr>
<tr>
<td>skipper</td>
<td>LTR</td>
<td>15–20</td>
<td>7</td>
<td>0.3</td>
<td>this work</td>
</tr>
</tbody>
</table>

aUnlike classical LTR elements, DIRS-1 lacksLTRs as well as recognizable protease and integrase domains.

### REFERENCES


### ACKNOWLEDGEMENTS

We thank Saskia van Es and Peter Devreotes for a gift of genomic DNA and Thomas Winckler and Theo Dingermann for the genomic library. Supported in part by NIH grants AI35282 to J.D.B. and GM47282 to T.L.S.

### The skipper element brings the total number of families of transposable elements isolated from *Dictyostelium* to five, and is the first true LTR element identified in this organism (Table 1). Together, the known transposons of this species are predicted to account for at least 3% of the total genomic DNA. The ability to transform *Dictyostelium* will allow the proposed unique modes of regulating the translation levels of its various gene products to be investigated.