The DIRS1 Group of Retrotransposons

Timothy J. D. Goodwin and Russell T. M. Poulter
Department of Biochemistry, University of Otago, Dunedin, New Zealand

Only three retrotransposons of the DIRS1 group have previously been described: DIRS1 from the slime mold Dictyostelium discoideum, PAT from the nematode Panagrellus redivivus, and Prt1 from the zygomycetous fungus Phycomyces blakesleeanus. Analyses of the reverse transcriptase sequences encoded by these elements suggest that they are related to the long terminal repeat (LTR) retroelements, such as the Ty3/gypsy retrotransposons and the vertebrate retroviruses. The DIRS1-group elements, however, have several unusual structural features which distinguish them from typical LTR elements: (1) they lack the capacity to encode DDE-type integrases or aspartic proteases; (2) they have open reading frames (ORFs) of unknown function; (3) they integrate without creating duplications of their target sites; and (4) although they are bordered by terminal repeats, these sequences differ from typical LTRs in that they are either inverted repeats or “split” direct repeats. Because of the small number of DIRS1-like elements described, and the unusual structures of these elements, little is known about their evolution, distribution, and replication mechanisms. Here, we report the identification of several new DIRS1-like retrotransposons, including elements from nematodes, sea urchins, fish, and amphibia. We also present evidence for the existence of DIRS1-like sequences in the human genome. In addition, we show that the lack of DDE-type integrase genes from elements of the DIRS1 group is explained by the finding that the previously uncharacterized ORFs of these elements encode proteins related to the site-specific recombinase of bacteriophage lambda. The presence of lambda-recombinase-like genes in DIRS1 elements also accounts for the lack of target-site duplications for these elements and may be related to the unusual structures of their terminal repeats.

Introduction

Long terminal repeat (LTR) retrotransposons are eukaryotic transposable elements that transpose via RNA intermediates. The integrated DNA forms of these elements are usually found flanked by LTR sequences. Active LTR retrotransposons encode the enzyme reverse transcriptase (RT), which is responsible for converting genomic RNA transcripts into DNA during the replication process. The RTs of all LTR retrotransposons are similar in sequence, suggesting that these elements share a common ancestor (Xiong and Eickbush 1990). Phylogenetic analyses based on alignments of RT sequences divide LTR retrotransposons into five major groups: the BEL, Ty1/copia, Ty3/gypsy, and DIRS1 groups and the vertebrate retroviruses (fig. 1A). The plant caulimoviruses and animal hepadnaviruses also encode RTs similar in sequence to LTR retrotransposons, although these elements are not LTR retrotransposons in the strict sense, as they have circular extrachromosomal DNA genomes and thus lack LTRs.

Many LTR retrotransposons have similar overall structures (fig. 1B). Most consist of two directly repeated LTRs flanking an internal region containing a small number of long open reading frames (ORFs). The ORFs usually include gag, which encodes proteins that form the structural component of a cytoplasmic particle within which the reverse transcription reaction takes place, and pol, which encodes several enzymes. In most elements, the Pol enzymes include an aspartic protease (Pro), an RT, a ribonuclease H (RNase H), and an integrase (Int). The Int genes of most LTR retrotransposons appear to be related to the transposase genes of certain classes of eukaryotic DNA transposons and bacterial insertion sequences, as the integrase/transposase proteins encoded by these elements each contain two highly conserved aspartate residues, followed about 35 residues downstream by a highly conserved glutamate residue (commonly known as the DDE motif), along with other sequence similarities (Capy et al. 1996, 1997). Some LTR elements also have a third ORF, env, which encodes envelope proteins that allow the elements to be transmitted between cells and between individuals, i.e., to be infectious. Env ORFs have most commonly been found in vertebrate retroviruses but have also been reported in members of the Ty3/gypsy (Kim et al. 1994; Song et al. 1994), Ty1/copia (Laten, Majumdar, and Gaucher 1998), and BEL (Malik, Henikoff, and Eickbush 2000; Frame, Cutfield, and Poulter 2001) groups.

Of the five groups of LTR retrotransposons, the one which is most poorly understood and which perhaps contains the elements with the most unusual structures is the DIRS1 group. Only three members of the DIRS1 group have previously been reported. These are DIRS1 itself from the slime mold Dictyostelium discoideum (Cappello, Handelsman, and Lodish 1985), PAT from the nematode Panagrellus redivivus (de Chastonay et al. 1992), and Prt1 from the zygomycetous fungus Phycomyces blakesleeanus (Ruiz-Perez, Murillo, and Torres-Martinez 1996). These elements all have structures quite distinct from those of typical LTR retrotransposons (fig. 1C). For instance, the termini of DIRS1 are inverted, rather than direct, repeats and are not delimited by the dinucleotides 5′-TG . . . CA-3′, unlike the LTRs of most LTR retrotransposons. The terminal repeats of DIRS1 are also not identical, with the right repeat having an additional 27-bp sequence (termed “re”) at its 3′ end.
that is not found in the left repeat. Furthermore, the 3' end of the element's internal region includes an 88-bp sequence, known as the internal complementary region (ICR), which is complementary to the outer edges of the element: the first 33 bp of the ICR are complementary to the start of the left terminal repeat, and the next 55 bp are complementary to the end of the right terminal repeat (including the 27-bp re section).

The termini of Prt1 are also inverted repeats and are, in addition, very short (~50 bp) compared with the LTRs of typical LTR retrotransposons (generally 200–1,000 bp). The termini of the PAT element are even more unusual than those of DIRS1 and Prt1 (fig. 1C). The sequence at the 5' end of the element is directly repeated in the internal region. The sequence from the 3' end of the element is also repeated in the internal region, immediately upstream of the internal copy of the 5' sequence.

The coding regions of the three DIRS1-type elements are also atypical (fig. 1C). DIRS1 contains three
long ORFs. The first of these is in an appropriate position and of an appropriate size to correspond to the gag ORFs of other retrotransposons. The second ORF (ORF2) overlaps the 3' end of the putative gag ORF and extends to near the 3' end of the element. No function has previously been assigned to ORF2. The third ORF is entirely overlapped by ORF2, but in a different reading frame, and encodes an RT and an RNase H. The element does not appear to encode an aspartic protease or a DDE-type integrase.

The ORFs of the sequenced copy of Prt1 are slightly corrupted by frameshifts and nonsense mutations, but a potential gag ORF and an ORF encoding an RT and an RNase H can be recognized. In addition, Prt1 contains a previously unreported third ORF. The PAT element also has a putative gag ORF, an ORF encoding RT and RNase H, and an uncharacterized third ORF. Like DIRS1, neither PAT nor Prt1 appears to encode an aspartic protease or a DDE-type integrase.

The three DIRS1-group elements have at least one further feature which distinguishes them from typical LTR retrotransposons: most other elements create short (4–6-bp) duplications of their target sites when they integrate, because the nicks produced by the integrase in each strand of the target site are slightly offset from each other. In contrast, the DIRS1-group elements all appear to integrate without creating such target-site duplications.

To improve our understanding of the DIRS1 group, we searched the large amounts of genomic DNA sequence currently available in the public databases for additional DIRS1-like elements, and we characterized the ORFs of all the known DIRS1-like elements. In this paper, we describe the detection and analysis of several new members of the DIRS1 group, including elements from a number of vertebrates, such as the zebrafish Danio rerio, the freshwater pufferfish Tetraodon nigroviridis, and the clawed toad Xenopus laevis. Furthermore, we show that the lack of DDE-type integrase genes in DIRS1 elements is explained by the finding that these elements all encode recombinases related to the site-specific recombinase of bacteriophage lambda. The presence of ι-recombinase-like genes in DIRS1 elements also accounts for the absence of target-site duplications for these elements and may be related to the unusual structures of their terminal repeats.

Materials and Methods

Sequence Analyses

Sequences were obtained from either the DNA Data Bank of Japan (DDBJ) database (http://www.ddbj.nig.ac.jp/) or the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/). BLAST searches were performed using either the DDBJ BLAST server (http://spiral.genomes.nig.ac.jp/homology/blast-e.shtml) or the NCBI BLAST server (http://www.ncbi.nlm.nih.gov/BLAST/). General sequence analyses were performed with the programs of the GCG package (Genetics Computer Group 1994). Multiple-sequence alignments were produced using CLUSTAL X (Thompson et al. 1997) and adjusted using SEAVIEW (Galtier, Gouy, and Gautier 1996). Phylogenetic trees were constructed using PHYLIP (Felsenstein 1989) and PAUP4.0 (Swofford 1998).

Accession Numbers

The accession numbers of many of the sequences described in this report are listed below. The accession numbers of the additional ι-recombinase sequences mentioned in figure 6 are listed in the legend to figure 6. Accession numbers are as follows: Agrobacterium tumefaciens tumor-Inducing plasmid, AF242881; bacteriophage lambda, J02459; bacteriophage P1, X03453; Caenorhabditis briggsae DIRS1-like sequences, AC090521, AC090839, AC084650, AC084491; Caenorhabditis elegans DIRS1-like sequences, AC090999, Z82079; Chlamydomonas reinhardtii transposon Pioneer1, U19367; Coxiella burnetii plasmid QpRS, Y15898; Danio rerio DIRS1-like sequences, AL590134, AL590155, AL591176, AL591172, AL591406, AL591588, AL591418, AL591144, AF112374, AL545740, AI958808, BG728155, G46488, G44922; Dictyostelium discoideum DIRS1, M11339; Fugu rubripes DIRS1-recombinase-like sequences, AL026865, AL125910, AL125934; Homo sapiens DIRS1-recombinase-like sequences, AC007833, AQ419451; Phycymyes blakesleeanus Prt1, Z54337; Panagrellus redivivus PAT, X60774; Pseudomonas transposon Tn5041, X98999; Saccharomyces cerevisiae 2 micron circle plasmid, J01347; Selenomonas ruminantium integrase, AB011029; Strongylocentrotus purpuratus DIRS1-like sequences, AZ192047, AZ187824, AZ157316, AZ157776, AZ181644, AZ145601, AZ173344, AZ145302, AZ137087 (and others); Xenopus laevis DIRS1-like sequences, BG555156, BG578087, BG364248, BG363884, BE576191, BG15648, BE575831, BG163190, AW460970 (and others); Xenopus tropicalis DIRS1-like sequence, BG514933.

The assembled sequence of the TnDirs1 element from the freshwater pufferfish T. nigroviridis, as well as the purple sea urchin S. purpuratus RT sequences, can be obtained from the authors’ website (http://biochem.otago.ac.nz:800/staff/poulter/pouliner.htm).

Notes on Terminology

The previous descriptions of DIRS1-like elements, together with the findings presented in this paper, suggest that the members of the DIRS1 group differ from all other known LTR retrotransposons in a number of important features. In particular, the terminal repeats of the DIRS1-like elements differ from typical “retroviral-type” LTRs in their structures and their probable mechanisms of action, and they may also have an independent origin. This creates the problem of whether the termini of DIRS1-group elements should be referred to as “LTRs” and whether DIRS1-like elements should be called “LTR retrotransposons.” Because of the close sequence similarity between the RT genes of DIRS1-group elements and typical LTR retrotransposons and the like-
lihood that their terminal repeats perform analogous functions (i.e., allowing the synthesis of a full-length cDNA from less-than-full-length RNA templates), in this paper we refer to DIRS1-like elements as LTR retrotransposons and their terminal repeats as LTRs. Where it is necessary to distinguish between the different types of LTRs, however, we suggest that the LTRs of DIRS1-like elements be referred to as DLTRs.

The new DIRS1-like elements described in this report have been given names beginning with two letters identifying the host species. For example, the Danio rerio elements are given names beginning “Dr.” The name of each full-length element then contains an indication of the previously identified element to which it is structurally most similar. For example, an element from Caenorhabditis briggsae that is similar in structure to PAT is named CbPat1. The names of elements whose full-length structures are not yet known consist of the species-identifying letters followed by the letter D and a number (e.g., DrD2) to indicate that they are DIRS1-related sequences, even though their overall structures are not known. Sequences containing just parts of DIRS1-like recombinase genes are given names consisting of species-identifying letters followed by “recom” and a number (for example, Hsrecom1).

Results
Identification of New DIRS1-like Elements

To identify previously uncharacterized DIRS1-like elements, we began by performing TBLASTN searches (protein sequence query vs. nucleotide sequence database) of the public sequence databases using the RT/RNase H sequences of the previously identified DIRS1-like elements as queries.

DIRS1-like Elements in Fish

In the nonredundant section of the databases, a high-quality match (E = 9 × 10^{-35} when using the DIRS1 RT/RNase H sequence as a query [where E represents the probability of a match of the observed quality occurring by chance]) was found to a sequence within the odorant receptor gene cluster of the zebrafish Danio rerio (bases 54803–56384 of the entry with accession number AF112374; Dugas and Ngai 2001). A reciprocal search, performed using the predicted RT/RNase H sequence of this zebrafish element as a query in a TBLASTN search of the databases, detected the three previously identified DIRS1-like elements as the top hits (apart from the zebrafish element itself), with E values ranging from 2 × 10^{-51} (DIRS1) to 1 × 10^{-24} (Prtl). The next best match, cauliflower mosaic virus, had an E value of 3 × 10^{-23}. These findings strongly suggest that this zebrafish sequence is a DIRS1-like retrotransposon. This element was named DrDirs1 (Danio rerio DIRS1-like element number 1; see Materials and Methods for naming conventions).

Additional copies of the DrDirs1 element were subsequently found in several zebrafish BAC sequences present in the High Throughput Genomic Sequence (HTGS) division of the public databases. These sequences were produced by a zebrafish genome sequencing project at the Sanger Centre (http://www.sanger.ac.uk/Projects/D.rerio/; accession numbers are listed in Materials and Methods). The structure of one of these DrDirs1 elements (AL590134) is illustrated in figure 2. It can be seen that this zebrafish element has an overall structure remarkably similar to that of DIRS1 itself. The element is ~6.1 kb in length, which is somewhat longer than DIRS1 (4.8 kb), but, like DIRS1, it is bordered by inverted repeats. These repeats are slightly different in sequence, with the right-hand copy having an additional short sequence at its 3’ end which is not present in the left copy. This sequence is 26–29 bp long, depending on how the exact termini of the element are defined (see below), and is thus very similar in length to the 27-bp re sequence found at the 3’ end of the right LTR of DIRS1 itself. The zebrafish DrDirs1 element also has a 97-bp ICR in the 3’ end of its internal region which contains adjacent sequences complementary to the outer edge of each LTR, similar to the 88-bp ICR of DIRS1.

The internal region of DrDirs1 contains several long ORFs. The first of these possibly corresponds to a gag ORF, as it is of an appropriate size (480 codons) and is located in a position similar to that of the gag ORFs of other retrotransposons, although no sequence similarity to previously identified gag ORFs was found. The second ORF encodes a putative protein bearing all of the expected highly conserved residues of RT and RNase H proteins. The DrDirs1 RT sequence is shown aligned with other RT sequences in figure 3. The 5’-two-thirds of the third ORF overlaps the majority of the RT/RNase H ORF but is in a different reading frame, similar to the way in which ORF2 of DIRS1 overlaps the DIRS1 RT/RNase H ORF. The putative protein sequence encoded by the 3’ end of this ORF is similar in sequence to the predicted product of the 3’ end of ORF2 of DIRS1 (see below), suggesting that these ORFs are homologous. Note that although the third ORF of the DrDirs1 element shown in figure 2 is disrupted by a single stop codon about halfway along its length, the corresponding positions in several other copies of the DrDirs1 element are, in contrast, occupied by glutamate codons. This suggests that this stop codon represents a nonsense mutation suffered by this particular copy of the element, rather than being a general feature of DrDirs1 elements.

The overall structures of the other sequenced copies of DrDirs1 are generally similar to that of the element shown in figure 2, except that each appears to contain several mutations which probably render them nonfunctional. For instance, the long DrDirs1 element present in the zebrafish odorant receptor gene cluster sequence (AF112374) has several nonsense and frameshift mutations, and its right LTR appears to have been deleted (not shown). It is also of interest to note that several expressed sequence tag (EST) sequences apparently derived from copies of DrDirs1 (AI545740, AJ958808, and BG728155) were identified, suggesting that some copies of DrDirs1 are transcriptionally active.

In addition to DrDirs1, a partial sequence of a second distinct family of zebrafish DIRS1-like elements
was identified. This element was named DrD2 and is present in sequence AL591418. The front half of this element appears to have been deleted, and the remainder, consisting of the 3’ half of the RT/RNase H gene and an additional gene (see below), has several frameshift and nonsense mutations (not shown).

In the Genome Survey Sequence (GSS) division of the public databases, we detected multiple high-quality matches to DIRS1-like RT/RNase H proteins (with E values ranging down to $5 \times 10^{-36}$ when using the DIRS1 sequence as a query, and down to $3 \times 10^{-74}$ when using the zebrafish DrDirs1 sequence as the query) in sequences from the freshwater pufferfish *T. nigroviridis*. These sequences were single reads from the ends of BAC, cosmid, and plasmid clones of *T. nigroviridis* DNA and were produced as part of the genome sequencing project currently active for this species (Roest Crollius et al. 2000). As for the zebrafish element described above, when reciprocal BLAST searches were performed using the *T. nigroviridis* sequences as queries, the top hits were the DIRS1-group elements, confirming that these sequences are derived from DIRS1-like elements.

Most of the *T. nigroviridis* sequences that we detected in these searches appear to belong to elements of one, apparently fairly abundant, family. This family was named TnDirs1. We were able to construct a full-length consensus sequence of TnDirs1 by identifying sequences containing overlapping fragments of the element and assembling these into a contig. The deduced structure of

---

**Fig. 2.**—Full-length DIRS1-group elements. The structures of all of the known full-length DIRS1-group elements are depicted. λ-recom. = lambda-recombinase (see the text). Note that the stop codon within the lambda-recombinase open reading frame of DrDirs1 appears to be a nonsense mutation suffered by only this particular copy of the element and that the *Tetraodon nigroviridis* sequence is a consensus sequence assembled from multiple random genomic sequence reads and, as a result, does not necessarily represent any one particular element in the genome and may contain errors.
the full-length element is illustrated in figure 2. The overall structure of TnDirs1 is remarkably similar to the structures of DIRS1 and DrDirs1: the element is \( \sim 5.9 \) kb in length and is bordered by inverted LTRs. The right LTR is slightly different from the left LTR, having an additional short (24–27-bp) sequence at its 3' end, and the internal region has a 91-bp ICR near its 3' end which contains sequences complementary to each end of the element.

Like DIRS1 and DrDirs1, the T. nigroviridis TnDirs1 element also contains three long ORFs. The first of these is possibly a gag ORF. Interestingly, the predicted product of this ORF is similar in sequence to that of the putative gag ORF of the zebrafish DrDirs1 element. The two proteins are 26% identical over a 471-amino-acid range, and they both have a region containing cysteine and histidine residues with similar spacing (\( \ldots Cx3Cx4Hx3Cxx3Hx5(CxHC) \)) near their N-termini which may form zinc fingers, which are commonly found in retrotransposon Gag proteins. (The putative DIRS1 Gag protein, in contrast, contains no obvious potential zinc fingers.)

The second ORF of TnDirs1 encodes a protein bearing all of the conserved domains of RT and RNase H proteins. The TnDirs1 RT sequence is shown aligned with other RTs in figure 3. The TnDirs1 RT/RNase H protein is 43% identical to the zebrafish DrDirs1 RT/RNase H protein over a 826-amino-acid range. It is a 27% match to the DIRS1 RT/RNase H protein over a slightly shorter (616-amino-acid) range. The third ORF of the T. nigroviridis element overlaps the RT/RNase H ORF for much of its length. The predicted product of the 3' portion of this ORF is similar in sequence to the predicted products of the 3' portion of ORF2 of DIRS1 and the 3' portion of the third ORF of the DrDirs1 element (see below), suggesting that these ORFs are homologous. Interestingly, the products of the 3' thirds of these ORFs from the zebrafish and T. nigroviridis elements are 50% identical in sequence (over 340 amino acids), showing that this part of each ORF has been at least as highly conserved as the RT/RNase H ORFs. The products of the 5' two-thirds of these ORFs (corresponding to the parts that overlap the RT/RNase H ORFs) are only 21% identical in sequence, showing that this part of each ORF has not been so highly conserved.

It is important to note that although the structures of the repeats of DIRS1, DrDirs1, and TnDirs1 are very similar, the actual sequences of these repeats are not
very similar in the different elements. This suggests that it is the ability of the repeats to form secondary structures, rather than their primary sequences, which is most important for the function of these elements.

**PAT-like Elements in Caenorhabditis**

Using the amino acid sequences of the PAT element as queries, we detected several copies of a PAT-like element in sequences from the nematode *C. briggsae* (finished cosmids sequences produced by a *C. briggsae* genome sequencing project at the Genome Sequencing Center at Washington University in Saint Louis: http://genome.wustl.edu/gsc/Projects/C.briggsae/). The newly identified element was named CbPat1. Two of the sequenced copies of CbPat1 appear to be extensively deleted, whereas a third copy (on sequence AC090521) seems to be relatively intact. The structure of this third copy is depicted in figure 2. The element is somewhat shorter than PAT itself (4.4 kb, compared with 5.5 kb) but otherwise has a very similar overall structure. Like PAT, it contains “split” direct repeats—the sequence from the 5' end of the element is repeated in the latter half of the internal region, and the sequence from the 3' end of the element is also repeated in the internal region, immediately upstream of the internal copy of the 5' end. The element also has three recognizable ORFs. The first of these is probably a *gag* ORF, as it is of an appropriate size and is in a position similar to those of other *gag* ORFs. Moreover, the predicted product of the ORF contains a putative Cx2Cx4Hx4C zinc finger, similar to those found in *Gag* proteins. A similar zinc finger is also encoded by the putative *gag* ORF of PAT itself (de Chastanay et al. 1992). The second ORF encodes an RT (fig. 3) and an RNase H. The third ORF encodes a protein similar in sequence to the third ORF of PAT and the additional ORFs of the other DIRS-group elements (see below). It is important to note that while PAT and CbPat1 are very similar in structure, the two elements are not highly similar in their actual sequences. For instance, the predicted products of their RT/RNase H ORFs are ~29% identical over a 580-amino-acid range, and their repeat sequences bear little resemblance to each other. This suggests that their “split-direct-repeat” structures have been maintained over a considerable period and are probably important for the replication of these elements.

**Additional DIRS1-Group Elements**

In addition to the largely intact fish and nematode elements described above, a variety of sequences containing fragments of new DIRS1-like elements from other species were identified using DIRS1-group RT/RNase H sequences as queries. For instance, in the GSS division, we detected many high-quality matches to DIRS1 elements in sequences from the purple sea urchin *S. purpuratus*, and in the EST division, several DIRS1-like elements from *X. laevis* and *X. tropicalis* were identified.

The new *S. purpuratus* DIRS1-like elements were found in sequences produced as part of a sea urchin genome sequencing project (Cameron et al. 2000). The elements we have identified fall into five different families, which were named SpD1–SpD5. Each element encodes a DIRS1-like RT bearing all the expected highly conserved residues. The RT sequences of SpD1–SpD4 are shown aligned with the other RTs in figure 3. There were insufficient sequences available for each of these elements to enable us to build up consensus sequences and thus to learn something about their structures. One of the elements (SpD2), however, has an uninterrupted ORF on the same strand as its RT/RNase H ORF, but in a different reading frame (not shown), raising the possibility that it has a structure similar to that of DIRS1 and the DrDirs1 and TnDirs1 elements described above. None of the other four elements appear to have overlapping ORFs, suggesting that these elements may have alternative structures.

The *Xenopus* sequences we identified included the *X. laevis* sequences with accession numbers BG555156, BG578087, BG364248, BG363884, BE576191, BG515648, and BE575831 and the *X. tropicalis* sequence BG514933. The *X. laevis* sequences usually differ from each other quite substantially in their overlapping regions, suggesting that *X. laevis* contains several distinct families of DIRS1-related elements. Only one of the *Xenopus* sequences (*X. tropicalis* BG514933) covers most of the highly conserved regions of RT. The element from which this sequence was derived is referred to as *Xid*1, and its RT sequence is shown aligned with other RT sequences in figure 3. As for the *S. purpuratus* elements, there was insufficient information available for each of the *Xenopus* elements to enable us to construct representative full-length sequences and thus to learn much about their structures. Most of the sequences, however, have ORFs overlapping their RT/RNase H ORFs on the same strand but in different reading frames. In some cases, sequence similarities between these ORFs and the ORFs overlapping the RT/RNase H ORFs in the DIRS1, DrDirs1, and TnDirs1 elements can be recognized (not shown), raising the possibility that these *Xenopus* elements may have DIRS1-like structures. Two of the *X. laevis* sequences (BE576191 and BG515648) appear to have uninterrupted RT/RNase ORFs, while lacking overlapping ORFs. These sequences are probably derived from elements with structures different from that of DIRS1. Interestingly, all of the *Xenopus* sequences were found in the EST database, suggesting that they may be derived from transcriptionally active elements.

**Phylogenetic Analyses**

An alignment of RT sequences similar to that shown in figure 3 was used to construct phylogenetic trees to examine the relationships among the various members of the DIRS1 group. The alignment contained representatives from each of the major groups of LTR retrotransposons and related viral groups, along with a non-LTR retrotransposon (*Tx1*) as an outgroup. An example of the trees obtained is shown in figure 4. The new DIRS1-like elements clearly group with the previously identified DIRS1-group elements, and this relationship is well supported by bootstrap resampling.
Within the DIRS1 group, the two full-length fish elements, DrDirs1 and TnDirs1, appear to be each other’s closest relatives. They also appear to be more closely related to DIRS1 than to either Prt1 or PAT, which is consistent with their structures being most similar to that of DIRS1. Similarly, PAT and CbPat1 are each other’s closest relatives and also have similar structures. Three of the four sea urchin elements group closely together, while the fourth represents a distinct lineage. The *Xenopus* element XtD1 groups most closely with the sea urchin element SpD2 and the two other vertebrate elements. None of the newly identified elements appear to be closely related to Prt1. The overall diversity of sequences within the DIRS1 group is comparable to that within the other major groups of LTR retrotransposons, despite the small number of DIRS1-like elements which have been identified.

The tree shown in figure 4 was constructed with the unweighted pair grouping method with arithmetic means (UPGMA) using PHYLIP (Felsenstein 1989). Similar results were also obtained, however, using PAUP (Swofford 1998) and neighbor-joining and parsimony methods. In all cases, the DIRS1-like elements appeared as a monophyletic group, and the relationships among the elements within the group were the same as, or very similar to, those shown in figure 4. The only major differences among the trees were that the branching order among the DIRS1, BEL, Ty1/copia, and hepadnavirus groups varied. For instance, in some trees, the BEL group appeared as the most basal branch, as in figure 4, whereas in other trees, the hepadnaviruses, or the hepadnaviruses and Ty1/copia elements together, formed the most basal branch. More detailed phylogenetic analyses will be required to determine the relative branching order of these major groups with any degree of certainty. On our trees, however, the DIRS1 group most frequently appeared as a sister taxon to a group composed of the caulimoviruses, the vertebrate retroviruses, and the Ty3/gypsy elements, as shown in figure 4.

**DIRS1 Elements Encode λ-Recombinases**

In an attempt to make sense of the unusual protein-coding capacities of DIRS1-group elements, we next sought to determine what the uncharacterized ORFs of these elements encode. Through a series of comparisons between the conceptual translation products of these ORFs and protein sequences present in the public sequence databases, we detected a convincing level of similarity between the products of these ORFs and a variety of recombinases related to the site-specific recombinase of bacteriophage lambda (fig. 5). We refer to
this class of sequences as λ-recombinases. In addition to the bacteriophage lambda recombinase, the λ-recombinase group includes, among other things, the recombinases of a number of other bacteriophages, the integrases or resolvases of some bacterial plasmids and transposons, the XerC and XerD recombinases of Escherichia coli which promote the stable inheritance of the E. coli chromosome, and the FLP-recombinases of yeast 2 micron circle plasmids (Argos et al. 1986; Hallet and Sherratt 1997; Nunes-Duby et al. 1998). The yeast FLP-recombinases, together with a single mitochondrial gene (Wolff et al. 1994) and a gene from an insect baculovirus (McLachlin and Miller 1994), are the only members of the λ-recombinase family previously identified in eukaryotes.

The λ-recombinases are very diverse in sequence. Only four highly conserved residues have been identified that are present in all, or nearly all, of the members of this family (Nunes-Duby et al. 1998). These are an apparently invariant Arg residue, followed about a hundred residues later by a highly conserved His residue, in turn followed two residues downstream by a second invariant Arg residue, which is followed 30 or so residues downstream by an invariant Tyr residue. These residues together are known as the RHRY tetrad and are believed to contribute to the active sites of the enzymes.

The previously uncharacterized ORFs of all the full-length DIRS1-group elements encode proteins bearing highly conserved RHRY tetrads similar to those of λ-recombinases (fig. 5). The highly conserved RHRY residues in the DIRS1-like elements are spaced similarly to those of known λ-recombinases, and sequence similarities between the DIRS1 elements and certain members of the λ-recombinase family are also evident in the regions surrounding the four conserved residues. The sequence similarities between the DIRS1-group elements and the previously identified λ-recombinases strongly suggests that the DIRS1-element proteins are members of the λ-recombinase class.

DIRS1-like Lambda-Recombinase Genes from Humans and Other Species

Using the putative DIRS1 element λ-recombinase sequences as queries, we conducted additional searches of the public sequence databases to see if any further DIRS1-group elements could be identified. We found eight previously undetected families of λ-recombinase sequences from S. purpuratus in the GSS database using DIRS1-group λ-recombinase sequences as queries. Some of these sequences may belong to the same elements as the five families of DIRS1-like RTs detected earlier in S. purpuratus. Four of these sequences cover the region containing the C-terminal part of the RHRY tetrad, which contains the conserved HRY residues. These sequences are shown aligned with the other λ-recombinases in figure 5B. We also found several DIRS1-recombinase-like sequences from the Japanese pufferfish F. rubripes in the GSS database (accession numbers AL125910, AL026865, and AL125934). These sequences are most similar to the T. nigroviridis element (not shown).

In the EST database, we detected 15 different X. laevis sequences carrying DIRS1-recombinase-like sequences. One of these (BG163190; Xrcom1) is shown aligned with the other λ-recombinases in figure 5B.

In the nonredundant database, we detected a second family of λ-recombinase from C. briggsae (on sequence AC084491) which is similar to, but distinct from, that of CbPat1, and we also detected a couple of sequences from C. elegans (Z82079 and AC090999) which also contain λ-recombinase-like genes similar to those of the C. briggsae sequences. One of these (Z82079; Cerecom1) is shown in figure 5B.

Interestingly, we also detected a DIRS1-like λ-recombinase gene in a sequence in the HTGS database (AC007833) which is annotated as being derived from H. sapiens. Using this putative human sequence as a query in further searches, we detected an almost identical, although shorter, sequence in the GSS database (AQ419451), which is also annotated as a human sequence. This putative human λ-recombinase is shown aligned with the DIRS1-recombinases and other λ-recombinases in figure 5B. The human λ-recombinase-like sequence is most similar to the DIRS1 recombinases, and, of these, it is most similar to those of the fish elements DrDirs1 and TnDirs1 and the sea-urchin element Sprecom1. These findings suggest that DIRS1-related sequences are present in the human genome. This human element is, however, probably nonfunctional, as it has suffered a frameshift mutation in its λ-recombinase ORF (not shown). Furthermore, no DIRS1-like RT or RNase H sequences were detected in the available human sequences, suggesting that if full-length DIRS1-like elements once existed in the human lineage, these elements must now be almost completely lost, or at least present in extremely low copy numbers.

One further point of interest to note is that while searching for DIRS1-like recombinase genes in eukaryotic sequences, we also identified a λ-recombinase-like gene in an unclassified transposon, Pioneer1 (Graham, Spanier, and Jarvik 1995), from the green alga C. reinhardtii. This transposon was first identified as a 2.8-kb insertion into an intron of the nitrate reductase gene of C. reinhardtii and appears to vary in copy number and genomic location in different strains (Graham, Spanier, and Jarvik 1995). The putative Pioneer1 recombinase sequence is shown in figure 5B. Although we could not unambiguously identify the first R of the RHRY tetrad in the Pioneer1 sequence, the HRY residues align well with λ-recombinases, and there are also sequence similarities in the regions flanking the highly conserved residues. This is the second report of a putative λ-recombinase gene in a plant, after a λ-recombinase-like gene in the Prototheca wickerhamii mitochondrion (Wolff et al. 1994), and raises the possibility that such sequences are more widespread in eukaryotes than is currently appreciated.

Relationships Among λ-Recombinases

The DIRS1-group recombinase sequences appear to be more similar to some members of the λ-recom-
**FIG. 5.**—Lambda-recombinase sequences in DIRS1 elements. A. A nearly full length alignment of the putative λ-recombinase sequences of DIRS1 elements and some bacterial λ-recombinase sequences. The four conserved residues of the RHRY tetrad (Argos et al. 1986; Nunes-
**DIRS1-like Retrotransposons 2077**

**Fig. 6.**—Relationships among lambda-recombinases. This tree is based on an alignment of the sequences encompassing the RHRY tetrads of a wide variety of λ-recombinases. It was constructed by the unweighted pair grouping method with arithmetic means using PHYLIP (Felsenstein 1989). The distance measure is that produced using the Categories option of the PROTDIST program. The host species and the accession numbers of the elements not listed in Materials and Methods are as follows: D29, mycobacteriophage D29, X70352; Ec, *Escherichia coli*, M31074; Ec-XerC, *E. coli*, X03923; Ec-XerD, *E. coli*, M54884; FLP-Kl, Klyveromyces lactis, X03961; Hi-rci, Haemophilus influenzae, U32821; P2, bacteriophage P2, AF063097; pSAM2, Streptomyces ambofaciens, X14899; SfbV, Shigella flexneri bacteriophage V, U82619; SF6, *S. flexneri* bacteriophage VI, X59553; SLP1, Streptomyces coelicolor plasmid SPL1, X71358; SsrA, Methanosarcina acetivorans plasmid pC2A, U78295; SSV1, Sulfolobus virus 1, X07234; T12int, Streptococcus pyogenes phage T12, U40453; Tn21, *E. coli* transposon Tn21, M33633; Tn4430, *Bacillus thuringiensis* transposon Tn4430, X07651; Tn554a and Tn554b, Staphylococcus aureus transposon Tn554, X03216; Tuc, bacteriophage Tuc2009, AF109874; XisA, *Anabaena* sp., U38537.

To investigate the relationships among the λ-recombinases in more detail, we constructed phylogenetic trees based on an alignment (not shown) of the sequences encompassing the RHRY tetrads of the available DIRS1-like λ-recombinases and a wide variety of other λ-recombinases. The full alignment contained 39 sequences and encompassed 207 amino acid positions. As noted earlier by Esposito and Scocca (1997), phylogenetic analyses of λ-recombinases are hindered by the great diversity of sequences within the family. Nevertheless, we found several features relevant to the evolution of DIRS1-like recombinase sequences which consistently appeared on trees constructed by different methods. These are illustrated in figure 6 by a tree obtained with UPGMA using PHYLIP. First, we found that the DIRS1-like recombinases always grouped closely with the four bacterial elements mentioned above, with
which they were most easily aligned, suggesting that the DIRS1 recombinases are indeed most closely related to these elements. In addition, the Cre recombinase of bacteriophage P1 (Sternberg et al. 1986) also consistently grouped with the DIRS1 recombinases. Second, the DIRS1-recombinases often fell into two groups, one consisting of the nematode PAT-like elements together with DrD2 and Xlrecom1, and the other comprising DIRS1, Prre, DrDirs1, TnDirs1, and Sprecom1. This division may be related to differences in the replication mechanisms of the elements, as all of the full-length elements in the PAT group that have been identified to date have split-direct-repeat structures, whereas all of the full-length elements in the DIRS1 group have inverted repeats. Third, the relationship of the DIRS1-like recombinases to the other major group of eukaryotic recombinases to the other major group of eukaryotic λ-recombinases, the yeast FLP-recombinases, is not clear. In some trees, such as the one shown in figure 6, the FLP-recombinases group separately from the DIRS1-like elements, whereas in others, they group with the DIRS1 sequences and the related bacterial elements. Clearly, more work will be required to resolve the evolutionary origins of the DIRS1-like recombinases, but at this stage, it appears that they are more closely related to a certain class of bacterial λ-recombinases, represented by the four sequences shown in figure 5A, than they are to other known λ-recombinases.

Integration of DIRS1-Group Elements

Given that DIRS1-group elements appear to encode proteins related to site-specific recombinases but do not encode DDE-type integrases, it is likely that the recombinases mediate the insertion of the putative extrachromosomal intermediates in the replication of these elements (Cappello, Handelsman, and Lodish 1985) into the host genome. Such a process would be unprecedented for retroelements. It is therefore of interest to examine the insertion sites of DIRS1-group elements to see what else can be learned about the integration of these elements.

The availability of multiple sequences allowed us to analyze the insertion sites of the zebrafish element DrDirs1, the pufferfish element TnDirs1, and the DIRS1 element itself in some detail. Our findings regarding the DrDirs1 element are illustrated in figure 7. Figure 7A depicts the sequences of the left and right termini of all the available DrDirs1 elements and their immediate flanking sequences. The regions that are highly conserved in all of the sequences are shown in boldface. It is evident that all of the sequenced copies of DrDirs1 are bordered at both their 5' and their 3' ends by GTT sequences and that there is little sequence similarity in the broader flanking regions. Figure 7B shows the sequence of eight sites which are closely similar in sequence to the regions flanking some of the DrDirs1 insertions shown in figure 7A but which lack a copy of DrDirs1 themselves. These sites are presumably similar in sequence to the occupied sites prior to the insertions of the DrDirs1 elements (and they represent all of the available unoccupied target sites that we could find).

These sites are shown aligned at the presumed points of insertion of the DrDirs1 elements. It can be seen that seven of these eight unoccupied target sites contain the sequence GTT at the point of insertion. The eighth has the sequence ATT at this site. Apart from this striking conservation of 3 bp at the insertion site, there is little else in the form of sequence similarities among the various target sites. These results suggest that DrDirs1 elements have a strong preference for insertion at GTT sequences.

A preference for insertion at GTT sequences in turn suggests a possible mechanism for integration of DrDirs1 elements (fig. 7C): if, as deduced by Cappello, Handelsman, and Lodish (1985), the final extrachromosomal replication intermediate of a DIRS1-like element is a circular molecule, then the junction between the ends of the DrDirs1 element would be likely to consist of the sequence GTT (underlined in fig. 7C). (This is suggested by comparison with the ICR, which, because it likely serves as the template for the synthesis of the proposed circular junction of the element’s termini [Cappello, Handelsman, and Lodish 1985], is probably identical in sequence to the circle junction.) The sequence similarity between the circular junction of the ends of the element and the sequence of the insertion site could then be used by the recombinase to perform a recombination reaction and thus insert the element into the host genome (fig. 7C). Such a process would result in the full-length element being flanked by GTT sequences. One of these repeats, or, more likely, part of one and the complementary part of the other, would be derived from the target site, while the remaining bases would be derived from the element. The uncertainty regarding which bases would be derived from the element and which would be derived from the target site creates the uncertainty regarding the exact lengths of the sequences alluded to earlier.

In support of this proposed mechanism of integration for DrDirs1, we found that the TnDirs1 elements are also flanked by GTT trinucleotides and appear to preferentially integrate at GTT sequences (data not shown). Likewise, from an alignment of 10 unoccupied DIRS1 target sites (not shown), we detected an apparent preference for DIRS1 insertions at sequences of the form A/T-T-T. (The composition of the three base positions in the 10 target sites is as follows: position 1, 4 T's, 5 A's; position 2, 8 T's, 1 A, 1 G; position 3, 10 T's.) The integration of DIRS1 could therefore occur by recombination between the putative circular junction of the element’s termini (AATT) and the target sites. Furthermore, the two sequenced copies of the PAT element (de Chastonay et al. 1992) are both flanked by AAC sequences, the internal repeat junction in the PAT element’s equivalent of the DIRS1 ICR (the internal copies of the element’s termini) consists of the sequence AAC, and the one sequenced copy of an unoccupied PAT target site also contains the sequence AAC at the insertion site. PAT may therefore insert by recombination between the AAC trinucleotide at the circle junction and an AAC sequence in the target site.
FIG. 7.—DrDirs1: target sites and a possible integration mechanism. 

A, The sequences at the 5' and 3' ends of all of the available DrDirs1 elements are shown aligned with each other. The regions that are conserved in most of these sequences, and which are probably directly related to the DrDirs1 insertions, are shown in boldface. The accession numbers of the sequences are indicated at the left. Note that sequence AL590155 contains two DrDirs1 3' ends.

B, Unoccupied DrDirs1 target sites. These sequences are very similar to the regions flanking some of the DrDirs1 elements in panel A but lack DrDirs1 elements themselves. Highly conserved trinucleotides at the insertion sites are shown in boldface. The accession numbers of the sequences containing the unoccupied sites are listed at the left. The sequences containing the corresponding DrDirs1 insertions are listed at the right. L = left; R = right.

C, A possible mechanism of DrDirs1 element insertion. The insertion of this element may occur by recombination between the GTT sequence at the circular junction of the element's termini and an identical sequence in the target site. The recombination reaction would be mediated by the element's encoded λ-like recombinase. The result would be an integrated element bordered by GTT sequences.

Overall, the available evidence is consistent with the possibility that the integration of a DIRS1-like element is mediated by the element's λ-like recombinase and suggests that the integration might occur by recombination between the 3-bp sequence at the circular junction of the element's termini and an identical sequence in the target site. Clearly, however, this proposed mechanism, while apparently plausible, will require experimental confirmation.

It should be noted that an obvious alternative mechanism for integration of DIRS1-group elements—that they integrate as linear molecules and create 3-bp duplications of their target sites—appears less satisfactory than the recombination mechanism outlined above, as it does not explain why the elements demonstrate such a strong preference for particular target sequences, it does not explain why this preference varies from element to element, and the
role played by the λ-recombinase in such a mechanism is not clear.

A final point of interest is that Cappello, Cohen, and Lodish (1984) noted that DIRS1 seems to preferentially insert into preexisting copies of itself, as five out of the six DIRS1 elements that were examined were located within another DIRS1-like sequence. We could find no evidence for a similar phenomenon associated with the additional elements examined here. For instance, of 10 distinct TnDirs1 termini examined, not one was inserted within or close to another TnDirs1 sequence, and of nine DrDirs1 termini examined, only one was within another DrDirs1 element.

Discussion

The identification of the new DIRS1-like elements described in this report extends the known host range of the DIRS1 group to include sea urchins, fish, amphibia, and mammals, whereas previously, they had only been found in a nematode, a slime mold, and a fungus. The DIRS1 group thus appears to be a fairly widespread collection of elements. No DIRS1-like elements have, however, been found in insects or plants, despite there being a large amount of sequence data available from these organisms.

Perhaps the most interesting aspect of this work is the discovery that all members of the DIRS1 group appear to encode a protein related to the recombinase of bacteriophage lambda and that they appear to lack the capacity to encode the more typical DDE-type integrases. The DIRS1 elements are unique among the known LTR retrotransposons in these features.

The presence of λ-recombinase genes in DIRS1 elements suggests that these elements insert into their host genomes by recombination, rather than by the DDE-family-integrase-mediated integration used by other LTR retrotransposons. This may provide an explanation for some of the unusual features of DIRS1-like elements. For instance, the apparent lack of target-site duplications associated with insertions of DIRS1-group elements (a feature restricted to LTR retrotransposons of the DIRS1 group) is neatly explained by integration employing a λ-recombinase, as the reactions catalyzed by these enzymes involve no synthesis or degradation of DNA (Hallet and Sherratt 1997).

The unusual LTRs of DIRS1 elements may also be related to integration mediated by a λ-recombinase: such a process would be facilitated by the preintegrative cDNA being circular in form, rather than linear (as is found in most other LTR retrotransposons). The unusual structures of DIRS1-element LTRs may be involved in the generation of circular full-length cDNAs. Interestingly, the final product of the reverse transcription reaction proposed for DIRS1 by Cappello, Handelsman, and Lodish (1985) in their original description of the DIRS1 element is a circular molecule.

It is as yet not clear where the λ-recombinases of the DIRS1-elements originated. The DIRS1-element recombinases are only the second major group of λ-recombinases identified in eukaryotes, after the FLP-recombinases of yeast 2 micron circle plasmids (although λ-recombinase-like genes have also previously been found in an insect baculovirus [McClelland and Miller 1994] and in a plant mitochondrial genome [Wolff et al. 1994], and in this work we identified an additional putative eukaryotic λ-recombinase in the Pioneer1 transposon from C. reinhardtii [Graham, Spanier, and Jarvik 1995]). Interestingly, however, the DIRS1-recombinases are more similar to a number of λ-recombinases from bacteria and bacteriophages than they are to the other known eukaryotic recombinases. Similarly, they are more similar to bacterial recombinases than to any known archaeal recombinases. Given this relationship, as well as the relative abundance of λ-recombinases in bacteria, it is possible that the DIRS1-recombinase genes are derived from a bacterial source.

It is also not clear how the ancestral DIRS1 RT gene came to be associated with a λ-recombinase gene. For instance, was this result of the acquisition of an LTR retrotransposon RT gene by a DNA transposon encoding a λ-recombinase? Or, perhaps less plausibly, was it the result of the replacement of a previously existing DDE-type integrase gene by a λ-recombinase gene in an ancestral LTR retrotransposon, together with a rearrangement of the element’s termini? Or was it the result of some other process? Further analyses of the relationships among the various groups of RT- and λ-recombinase-encoding sequences, as well as characterizations of additional elements, may help answer this question.

Analyses of the termini and unoccupied target sites of various DIRS1-like elements suggested that these elements insert into their host genomes by recombinations between particular 3-bp sequences at the circular junctions of the elements’ termini and identical 3-bp sequences in the target sites. It is probable, however, that there are additional sequences within the elements themselves that are required for the recombination reactions. This is suggested by the way in which the elements appear to use only the particular 3-bp sequences at the circular junctions of their termini as substrates, rather than identical 3-bp sequences at other locations within the elements. On the other hand, the only similarities among the various target sites of each element that we could identify were the conserved 3-bp sequences at the immediate insertion sites. If, as appears to be the case, these 3-bp sequences are the only sequence requirements in the target sites, then this represents a departure from the sequence requirements of well-characterized λ-recombinases. For instance, the recombinase of bacteriophage lambda itself acts at a 15-bp sequence common to the phage genome and the E. coli chromosome (Mizuuchi and Mizuuchi 1980), and the Cre recombinase of bacteriophage P1 acts at a particular 34-bp sequence in the phage genome (Hoess, Ziese, and Sternberg 1982). This apparent difference in sequence requirements at the target sites may reflect differences in the life cycles of the elements. For instance, it may be advantageous for DIRS1-like elements to have multiple potential insertion sites in the host genome, and disadvantageous for the reverse reaction (excision of the element) to occur. This could lead to selection for a short
target sequence and for most of the sequences controlling the specificity of the reaction to lie within the element. Unfortunately, not one of the bacterial recombinases to which the DIRS1-type recombinases appear to be most closely related (represented by the four elements in fig. 5A) has been functionally characterized, thus preventing a comparison with the sequence requirements of these recombinases.

While it has become apparent that the 3’ end of ORF2 in DIRS1 and the 3’ ends of the homologous ORFs in the DrDirs1 and TnDirs1 elements (and also the additional ORFs in PAT, CbPat1, Prt1, etc.) encode λ-recombinases, the question of what the 5’ two-thirds of these ORFs encode remains unanswered. On the basis of the findings that (1) the 5’ two-thirds of these ORFs overlap the highly conserved RT/RNase H ORFs, (2) that the predicted products of these sections of these ORFs are not well conserved among the different elements, and (3) that corresponding additional coding regions appear to be absent from PAT, CbPat1, and Prt1, we propose that the amino acid sequences encoded by these regions are not important at all. Rather, it is simply the presence of uninterrupted ORFs that is critical. The reasoning behind this proposal is that such overlapping ORFs would provide an alternative to mRNA splicing (or the generation of additional transcripts), to allow the translation of the downstream λ-recombinase-coding region, in such a way that the λ-recombinase would not end up being covalently attached to the RT/RNase H protein. The importance of having separate RT/RNase H and recombinase proteins is suggested by the way in which more typical LTR retrotransposons generally cleave their primary pol translation products, containing covalently attached RT/RNase H and Int proteins, into their constituent domains in order to activate those domains. Proteolytic cleavages of polypeptide precursors might not be possible for DIRS1-group elements, as no protease genes have so far been detected in these elements.

The TnDirs1 element from the freshwater pufferfish T. nigroviridis, and the DrDirs1 element from the zebrafish D. rerio were found to share some very unusual structural features with DIRS1 itself: (1) the three elements all contain inverted terminal repeats; (2) the left and right repeats of each element are slightly different in sequence, with the right LTRs having additional short sequences at their 3’ ends which are not found in the left LTRs; and (3) the 3’ end of the internal region of each element contains a region, known as the ICR, which consists of sequences complementary to the outer edges of each LTR. These specific features are not, however, similar in sequence in the different elements, nor are the elements in general highly similar in sequence. The conservation of these unusual structures in elements which have diverged considerably in sequence supports the proposal of Cappello, Handelsman, and Lodish (1985) that these structures are critical in the replication cycle.

It is of interest to note that PAT and CbPat1 are also very similar in structure, despite having diverged considerably in sequence. This suggests that the split-direct-repeat structures of these elements are critical for their replication. Finally, it is worth noting that even though the PAT-like elements differ considerably in structure from the DIRS1-like elements, these elements all share an unusual structural feature—copies of their terminal sequences present adjacent to each other within their internal regions. This similarity suggests that some features of the replication cycles may be similar in the two types of elements, despite the PAT-like elements having direct repeats and the DIRS1-like elements having inverted repeats. In contrast, the Prt1 element appears to lack internal copies of its terminal sequences, suggesting that the replication cycle of this element may differ in some features from the replication cycles of DIRS1 and PAT.

The findings presented here show that the DIRS1 group of LTR retrotransposons, which has received little attention to date, is a widespread and very interesting class of elements. While this study has clarified a number of features of the structure and evolution of these elements, many important questions remain. For instance, how exactly do these elements replicate? What was the source of the λ-recombinase gene? How did an LTR retrotransposon-like RT gene become associated with a λ-recombinase gene? We hope that this report will stimulate further research into these and other questions.

**Acknowledgments**

Most of the zebrafish sequences analyzed in this report were produced by the Danio rerio Sequencing Project at the Sanger Centre (http://www.sanger.ac.uk/Projects/D_rerio/). The C. briggsae sequences were produced by the Genome Sequencing Center at Washington University in Saint Louis (http://genome.wustl.edu/gsc/Projects/C.briggsae/). Most of the Xenopus sequences were produced by the Xenopus EST project, also at the Washington University Genome Sequencing Center (http://genome.wustl.edu/est/xenopus_estmpg.html).

**LITERATURE CITED**


Thomas Eickbush, reviewing editor

Accepted July 16, 2001