kangaroo, a Mobile Element From *Volvox carteri*, Is a Member of a Newly Recognized Third Class of Retrotransposons

Leonard Duncan,[1] Kristine Bouckaert, Fay Yeh[2] and David L. Kirk

Department of Biology, Washington University, Saint Louis, Missouri 63130

Manuscript received February 28, 2002
Accepted for publication August 30, 2002

ABSTRACT

Retrotransposons play an important role in the evolution of genomic structure and function. Here we report on the characterization of a novel retrotransposon called *kangaroo* from the multicellular green alga, *Volvox carteri*. *kangaroo* elements are highly mobile and their expression is developmentally regulated. They probably integrate via double-stranded, closed-circle DNA intermediates through the action of an encoded recombinase related to the λ-site-specific integrase. Phylogenetic analysis indicates that *kangaroo* elements are closely related to other unorthodox retrotransposons including *PAT* (from a nematode), *DIRS-1* (from Dictyostelium), and *DrDIRS1* (from zebrafish). *PAT* and *kangaroo* both contain split direct repeat (SDR) termini, and here we show that *DIRS-1* and *DrDIRS1* elements contain terminal features structurally related to SDRs. Thus, these mobile elements appear to define a third class of retrotransposons (the DIRS1 group) that are unified by common structural features, genes, and integration mechanisms, all of which differ from those of LTR and conventional non-LTR retrotransposons.

RETROTRANSPOSONS are mobile genetic elements that are found in a wide range of eukaryotes (Xiong and Eickbush 1990; Gabriel and Boeke 1993). They use reverse transcriptase (RT) to convert RNA intermediates into DNA copies that can then be integrated in new locations. Such replicative transposition means that retrotransposons can greatly influence genome size: it is estimated that ~40% of mammalian genomes and ~50% of the maize genome is composed of retroelements (SanMiguel *et al.* 1996; Smit 1999). In addition to increasing genome size, retroelements can affect genome structure and function in other ways. Although they often inactivate the gene into which they insert, there are now many examples in which novel *cis*-regulatory sequences or protein domains are believed to have been acquired from retroelements (Kumar and Bennetzen 1999; Smit 1999). For example, telomerase may have obtained its RT function this way (Eickbush 1997). Retroelements can also transduce flanking sequences, alter splicing of chimeric pre-mRNAs, create pseudogenes, and promote unequal crossing over and other genomic rearrangements (Finnegan 1989; Coffin 1993; Kumar and Bennetzen 1999).

Most retrotransposons can readily be placed in either the long terminal repeat (LTR) or the non-LTR classes (Figure 1A; Xiong and Eickbush 1990; Gabriel and Boeke 1993). LTR retrotransposons are closely related to retroviruses and are bounded by direct repeats that contain transcription-initiation and polyadenylation signals. They typically contain one open reading frame (ORF) that encodes a nucleic acid binding protein (Gag) and a second ORF that encodes protease, RT, RNAse H, and integrase domains (Figure 1A). They sometimes contain a third ORF encoding an envelope protein. Non-LTR elements are simpler than (and probably ancestral to) the LTR class (Xiong and Eickbush 1990): they lack terminal repeats, typically contain a poly(A)-rich sequence near their 3′ ends, and usually contain one ORF encoding a Gag protein and a second ORF encoding endonuclease, RT, and RNAse H functions (Figure 1A).

LTR and non-LTR elements use distinct transposition mechanisms (Craig 1997). LTR elements use RT to generate a free, linear cDNA copy of the element that is inserted into a target site by the action of an integrase related to DNA transposases. In contrast, non-LTR elements transpose by "target-primed reverse transcription," in which an RNA copy of the element is reverse transcribed only after an encoded endonuclease has cleaved the target DNA to generate a primer for reverse transcription.

RT-based molecular phylogenies generally identify retroelement clades whose individual members share other important features (Doolittle *et al.* 1989; Xiong and Eickbush 1990; McClure 1993). For example, the elements placed in the *gypsy*, *copia*, *BEL*, and retrovirus clades on the basis of their RT sequence all contain LTRs. Furthermore, elements in the *copia* clade all have an integrase gene upstream of the RT gene, whereas in

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession no. AY137241.

[1]*Corresponding author:* Cumbre, Inc., 1502 Viceroy Dr., Dallas, TX 75235. E-mail: len.duncan@cumbre.net

[2]*Present address:* University of Connecticut School of Medicine, 263 Farmington Ave., Farmington, CT 06030.

other LTR lineages (such as *gypsy*) the integrase gene is downstream of the RT gene.

A notable exception to this rule is the basis of this report. An RT-based phylogeny places two unorthodox retrotransposons—*PAT*, from the nematode *Panagrellus redivivus* (de Chastonay *et al.* 1992) and *DIRS-1* from *Dictyostelium discoideum* (Cappello *et al.* 1985)—as nearest neighbors on a branch located near the *copia* and *gypsy* clades (Malik and Eickbush 2001). However, neither *PAT* nor *DIRS-1* encode either an LTR type of integrase or a non-LTR type of endonuclease. Moreover, *PAT* and *DIRS-1* possess termini that appear to be very different from one another and from the termini of any other retroelement group. *PAT* contains split direct repeat (SDR) termini, in which one copy of an ~300-bp direct repeat is found in the interior of the element (juxtaposed open and solid triangles, Figure 1B), while the second copy is bifurcated, with about one-half of it present at each terminus (solo solid and open triangles, Figure 1B), such that the half-repeats alternate in the order A, BA, B. *DIRS-1*, on the other hand, contains inverted terminal repeats (Figure 1B).

Our efforts to develop transposon-tagging tools for use in studying the developmental genetics of the multicellular green alga, *Volvox carteri* (Kirk 1998), led us to the discovery of a highly mobile retroelement called *kangaroo* that contains SDR termini like those of *PAT* and that is closely related to *PAT* and *DIRS-1* in terms of its RT sequence. Here we report on the characterization of *kangaroo,* its developmentally regulated expression, and its probable method of integration.

The unusual features shared by *PAT*, *DIRS-1*, and *kangaroo-1* suggest that these elements must transpose by a mechanism distinct from that used by either LTR or non-LTR retroelements, despite the similarity of their RT proteins to those of well-known LTR elements. Because similar unorthodox elements are also found in the genomes of zebrafish and other metazoans, we conclude that this DIRS1 group represents a widespread third class of retrotransposons.

MATERIALS AND METHODS

Volvox strains and cultivation conditions: Strains HK 9 (male) and HK 10 (female) of *V. carteri* f. *nagariensis* were isolated and described by Starr (1969) and later provided to us by the University of Texas Culture Collection of Algae. EVE is our standard subclone of HK 10 (Harper *et al.* 1987) and CRH7 (Miller *et al.* 1993) and LDV45 (L. Duncan, unpublished data) are first- and second-generation subclones of EVE. The NIES male and female strains were isolated in 1983 from a slightly different area of Japan than HK 9 and HK 10 had been and were provided to us by the National Institute for Environmental Studies in Ibaraki, Japan. Volvox cultures were maintained in standard Volvox medium under standardized culture conditions (Kirk and Kirk 1983) on a 16-hr-light/8-hr-dark cycle.

Nucleic acid purification and sequencing: Volvox genomic DNA was purified as described (Miller and Kirk 1999). Plasmid DNA, λDNA, and hybridization probes were isolated using purification kits from QIAGEN (Valencia, CA). Volvox RNA was purified by minor modifications of a previously described method (Kirk and Kirk 1985). DNA was sequenced with Applied Biosystem's (Foster City, CA) BigDye v. 2.0 premix using standard methods. Sequence data were collected on a departmental MJ Research (Waltham, MA) Basestation and analyzed using PHRED/PHRAP/CONSED (http://www.phrap.org).

Cloning *kangaroo-1*: Two novel *Hin*dIII fragments that resulted from the insertion of *kangaroo-1* into the *nitA* (nitrate reductase-encoding) gene of CRH7 were separately cloned from subgenomic libraries generated from size-selected *Hin*dIII fragments of CRH7 genomic DNA ligated to *Hin*dIII-digested pBluescript II KS. The resulting 3.8-kb (pLD41) and 7.5-kb (pLD40) inserts containing two segments of *kangaroo-1* were sequenced using a combination of primer walking and the Genome Priming System (NEB) to a final estimated error rate of <0.01/10,000 bp. Next, we used OLV32 (5′-TTGTT GGGCTGCTTTCCTC-3′) and OLV46 (5′-GGAAGCACACGAA GTTGG-3′; Figure 2A) as PCR primers to amplify from CRH7 genomic DNA a 2.7-kb DNA fragment that spans the internal *Hin*dIII site within *kangaroo-1*. The sequence of this fragment confirmed that these two *Hin*dIII fragments that contain *kangaroo-1* sequences are juxtaposed as shown in Figure 2A.

Nucleic acid hybridization: Southern and Northern blotting experiments followed standard procedures (Sambrook *et al.* 1989). [α-^{32}P]dCTP-labeled probes were prepared using the oligolabeling kit (Pharmacia, Piscataway, NJ) and purified using NucTrap columns (Stratagene, La Jolla, CA).

DNA fragments used as hybridization probes in the figures were as follows: Probe 2 was a 1.3-kb *Sma*I fragment isolated from pLD41. Probe 3 was an ~600-bp fragment generated by PCR amplification from pLD40 using oligonucleotides OLV25 (5′-GCACTTACGACCGTGAAACC-3′) and OLV67 (5′-AAAACG GACGCTCCACGA-3′). Probe 4 was an ~4-kb *Xmn*I-*Hin*dIII fragment isolated from pLD40. Probe 5 was an ~1-kb fragment generated by PCR amplification from pLD40 using oligonucleotides OLV22 (5′-ATCCATCTTCGTATTTGCTG-3′) and OLV23 (5′-ACGAACGGGAGCACACTTAT-3′). Probe 6 was an ~1.4-kb *Hin*dIII-*Dra*I fragment isolated from pLD40. Probe 7 was a 325-bp fragment generated by PCR amplification from pLD41 using oligonucleotides OLV50 (5′-AATAGCGGGAAAGGG ATG-3′) and OLV63 (5′-GAAGTGTGAAGCCGACGA-3′). The C38 probe has been described previously (Tam and Kirk 1991).

Isolation of *kangaroo-2* termini and preinsertion site: *kangaroo-2* was identified by Southern blot analysis (using probe 2, Figure 2A) as a 3.5-kb *Bam*HI restriction fragment length polymorphism (RFLP) present in LDV45 but absent from EVE. The DNA fragment corresponding to this RFLP was cloned, generating pLD35. An ~400-bp fragment of DNA (probe 8) derived from the nonretrotransposon sequence that flanks the left side of *kangaroo-2* was amplified from pLD35 by PCR using oligonucleotides OLV9 (5′-ATGGATGGGACTT GCTGCTGAC-3′) and OLV10 (5′-CACCAATTTACCCGCCA GGATG-3′). Southern blotting experiments demonstrated that probe 8 hybridized with a single copy sequence in LDV45 and EVE genomic DNA and hybridized to the same 3.5-kb *Bam*HI RFLP present in LDV45 that is recognized by probe 2.

To isolate the *kangaroo-2* preinsertion site, probe 8 was used to screen an EVE genomic library constructed in λDASH II (Kirk *et al.* 1999). The preinsertion site was sequenced directly from bacteriophage DNA isolated from three independent probe 8-hybridizing clones.

Next, the sequence of the preinsertion site was used to design an oligonucleotide, OLV58 (5′-CACAGGGCGGGCAG TTAT-3′), whose sequence was expected to be present within the nonretrotransposon DNA that flanked the right side of

kangaroo-2. Two *kangaroo* specific oligonucleotides, OLV42 (5′-AGATTTGAGGCAGAGTAGG-3′) and OLV43 (5′-AGAAGA CACAGTCGGATGAG-3′), were separately used in conjunction with OLV58 to PCR amplify an ~1-kb fragment from LDV45 genomic DNA containing the junction between the right side of *kangaroo-2* and flanking DNA. Both independent PCR products were sequenced.

Sequencing of retrotransposon:flanking DNA junctions from *kangaroo-3* through *kangaroo-13*: Plasmids pLD33 and pLD34 were obtained by screening a partial LDV45 genomic library with *kangaroo* probe 2, and they contain distinct inserts that include the left termini of *kangaroo-5* and *kangaroo-4*, respectively. *kangaroo-3* and *kangaroo-6* through *kangaroo-13* were isolated by screening an EVE λ genomic library (Kirk *et al.* 1999) with a labeled 2.8-kb *Xmn*I DNA fragment of *kangaroo-1* derived from pLD40. The *kangaroo*:flanking DNA boundary sequences for several of these clones were determined by directly sequencing the corresponding purified bacteriophage DNA with oligonucleotides that hybridize to the left or right terminus of *kangaroo* elements. For *kangaroo-3* and *kangaroo-6*, however, *Sal*I DNA fragments containing the right side of the retroelement plus associated flanking DNA were first subcloned into pBluescript II SK to generate pLD43 and pLD42.

PCR amplification of a portion of the putative circular form of *kangaroo*: PCR reactions containing 0.5 μM OLV2 (5′-AAG ACACAGTCGGATGAGGAG-3′), 0.5 μM OLV93 (5′-CATTCT GGTGTCCTCCTT-3′; Figure 7B), and 0.5 μg of EVE DNA were carried out using standard methods (Sambrook *et al.* 1989). The predominant PCR product, which was the expected size (850 bp) to have been generated from a circular form of *kangaroo*, was cloned into pGEM-T Easy (Promega, Madison, WI) to generate pLD53, which was then sequenced.

Isolation of *kangaroo*-hybridizing cDNA clones: We purified several clones that hybridized with *kangaroo* probe 4 (Figure 6A) from EVE cDNA libraries constructed in λgt10 (Tam and Kirk 1991; clones λ5-11 and λ5-13) or λ-Uni-ZAP XR (B. Taillon and D. Kirk, unpublished data; clones λ5-2, λ5-4, and λ5-5). The inserts in λ5-2, λ5-4, and λ5-5 were converted to phagemids using the Rapid Excision kit (Stratagene) to generate pLD48, pLD49, and pLD50, respectively. The inserts from λ5-11 and λ5-13 were PCR amplified from purified bacteriophage DNA using λgt10 forward and reverse oligonucleotides and cloned into pGEM-T Easy to generate pLD56 and pLD57, respectively.

RESULTS

Isolation of *kangaroo-1*: Miller *et al.* (1993) previously used a selection for chlorate-resistant individuals to enrich for *V. carteri* mutants carrying transposon insertions in the nitrate-reductase-encoding gene, *nitA* (Gruber *et al.* 1992). One such mutant, CRH7, was found to contain a large insertion within *nitA* that was unrelated to the transposon that became the major focus of that study (Miller *et al.* 1993). We confirmed the existence of such an insertion element by probing a Southern blot containing *Hin*dIII-restricted genomic DNAs with a probe derived from the 5′ end of the *nitA* coding region: The ~2.5-kb *Hin*dIII fragment derived from the wild-type *nitA* gene was replaced in CRH7 by 3.8- and 7.5-kb restriction fragments (data not shown), a result consistent with the insertion of an ~9-kb DNA element. We cloned and sequenced both novel *Hin*dIII fragments, thereby establishing the structure of the inserted element (Figure 2A), which we call *kangaroo-1*. We then confirmed (as described in materials and methods) that the two *Hin*dIII fragments are juxtaposed as shown.

***kangaroo-1* is an unorthodox retrotransposon:** *kangaroo-1* possesses termini (solo solid and open triangles, Figure 2A) that are distinct from those of DNA transposons and most retrotransposons and that are similar in structure (but not in sequence) to the SDRs found in the *PAT* retroelement (Figure 1B). The terminal half-repeats of *kangaroo-1* are identical in sequence (data not shown) to their counterparts within the full-length internal repeat (juxtaposed open and solid triangles, Figure 2A). Near its left end, *kangaroo-1* also contains ~12 contiguous copies of an 89-bp sequence (Figure 2A). The first 10 copies of the 89-bp repeat are extremely similar, although many of them can be distinguished by a small number of nucleotide deletions and/or polymorphisms (Figure 2B). The last two repeats (not shown) are less well conserved and more difficult to align.

The two largest uninterrupted ORFs predicted by the nucleotide sequence of *kangaroo-1* are shown in Figure 2A. ORF-A and ORF-B potentially encode proteins 417 and 829 amino acids long, respectively. Importantly, a reverse position-specific BLAST search revealed that a portion of the deduced amino acid sequence of ORF-B is strikingly similar ($E = 6e\text{-}13$) to the RT family of proteins in the Pfam database (Bateman *et al.* 2002). This similarity extends over and includes the conserved RT regions 2–7 as defined by Xiong and Eickbush (1990) and the key conserved residues of RNAse H (Doolittle *et al.* 1989; McClure 1993; Figure 3). Region 5 of RT includes a highly conserved (Y/F)XDD box that is thought to be essential for binding divalent metal ions (Kohlstaedt *et al.* 1993). Although region 5 of RT from *kangaroo-1* and two other independent *kangaroo* clones (data not shown) encodes a less common LIDD (solid triangle, Figure 3), such divergence has been observed previously in other RT and related proteins (Doolittle *et al.* 1989; Xiong and Eickbush 1990). We therefore conclude that *kangaroo-1* encodes RT/RNAse H and is likely to be a retrotransposon. We note that in terms of nucleotide sequence, predicted amino acid sequences, overall organization, and phylogenetic position, *kangaroo-1* is clearly distinct from members of the *copia* class of retrotransposons (*e.g.*, *Osser*) that are found in *V. carteri* (Lindauer *et al.* 1993).

Interestingly, the RT protein from *kangaroo-1* is most closely related to the RT proteins encoded by *PAT* (BLASTP score: $E = 4e\text{-}24$) and *DIRS-1* (BLASTP score: $E = 6e\text{-}16$). Furthermore, as we discovered through TBLASTN searches of the databases and as was reported by Goodwin and Poulter (2001), the *Danio rerio* (zebrafish) genome contains several copies of a 6.1-kb retrotransposon (*DrDIRS1*) that is strikingly similar in structure to *DIRS-1* (Figure 1B) and encodes a deduced RT protein very similar to that of *kangaroo-1* (BLASTP score: $E = e\text{-}16$). An alignment of a portion of the RT/RNAse H

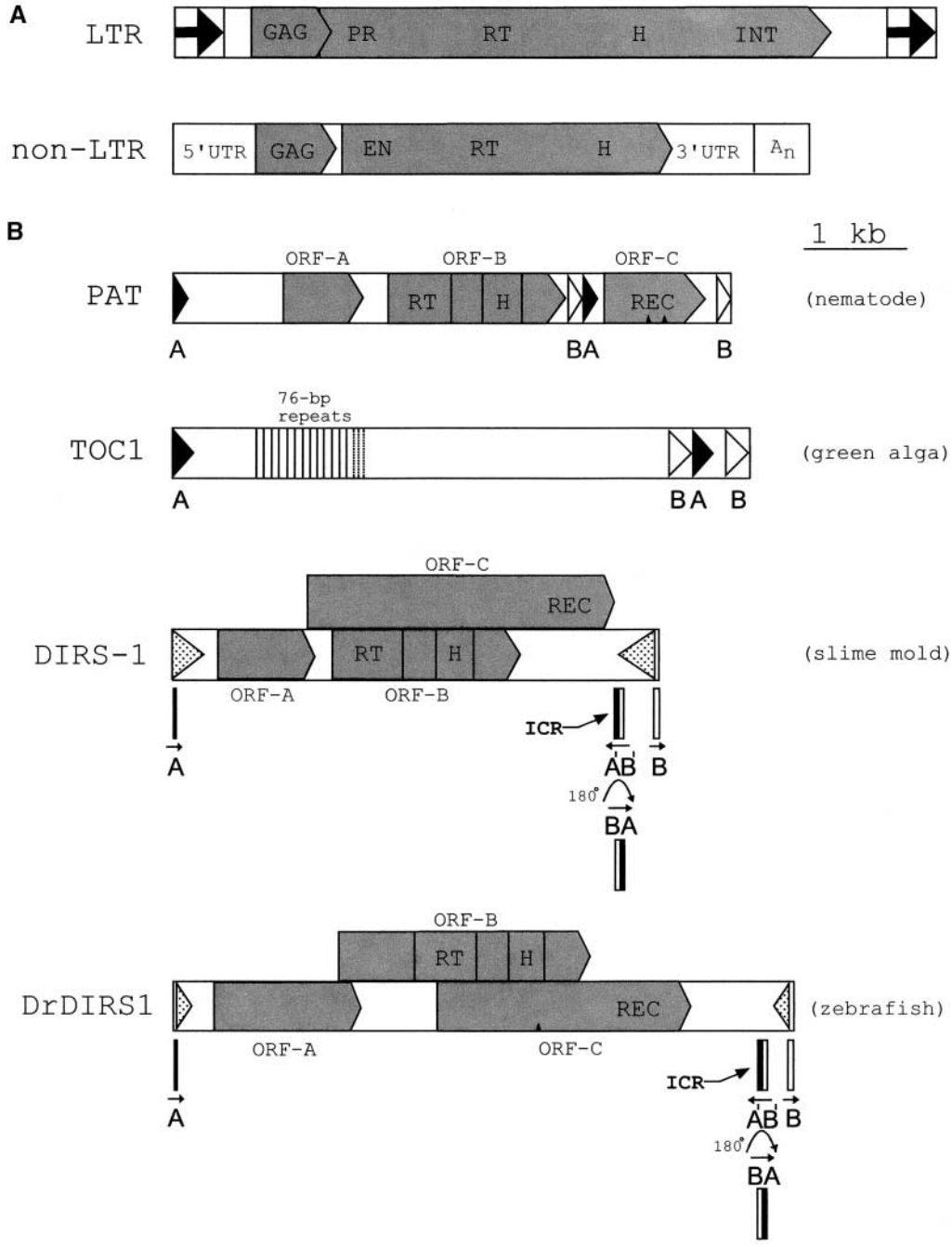

FIGURE 1.—Three major groups of retrotransposons. (A) The generic structures of LTR and conventional non-LTR retroelements. The figures are not to scale and are not intended to represent specific retrotransposons. Large black arrows, long terminal repeats; PR, protease domain; RT, reverse transcriptase domain; H, RNAse H domain; INT, integrase domain; UTR, untranslated region; EN, endonuclease; A_n, poly(A)-rich sequence. Major ORFs are indicated by shading. (B) Diagrams of four members of the DIRS1 group of retroelements. *PAT* and *TOC1* contain SDR termini (solid and open triangles designated A and B), while the termini of *DIRS-1* and *DrDIRS1* contain inverted repeats (dotted triangles). *DIRS-1* and *DrDIRS1* also contain short terminal sequences (solid and open rectangles labeled A and B) that overlap with part of and, in some cases, extend beyond the inverted repeats and that are repeated in an internal complementary region (ICR), where they are present in a juxtaposed and inverted arrangement (A′B′). Inversion of the ICR would create a structure very similar to that found in the SDR elements. When comparing any two DIRS1-group elements, the A and B repeat units are similar in structure but unrelated in sequence. REC, recombinase. Accession numbers for *PAT*, *DIRS-1*, and *TOC1* are X60774, M11339, and X56231, respectively. The copy of *DrDIRS1* shown here is found in accession no. AL590134. The specific *PAT* and *DrDIRS1* elements shown here contain nonsense codons (small solid triangles) in their *rec* genes. A second copy of *DrDIRS1* (not shown) and, presumably, other copies of *PAT* contain an uninterrupted ORF-C.

domains from these four related retroelements is shown in Figure 3.

When we used PAUP* 4.0 to perform a neighbor-joining phylogenetic analysis of the RT domains of *kangaroo-1* and 22 other RT proteins, a tree (not shown) was produced that was very similar in overall topology to those previously published (XIONG and EICKBUSH 1990; MCCLURE 1993; MALIK and EICKBUSH 2001). This phylogram indicated (with a bootstrap value of 77%) that the RT protein sequences from *kangaroo-1*, *PAT*, *DIRS-1*, and *DrDIRS1* constitute a clade—the DIRS1 group—that apparently diverged after the *copia* group and at about the same time as the retrovirus and *gypsy* groups. This finding suggests that although the DIRS1 group of retrotransposons includes members with two different types of unusual termini, these subfamilies may nevertheless share a common mode of replication. In this regard, it is particularly noteworthy that the similar-

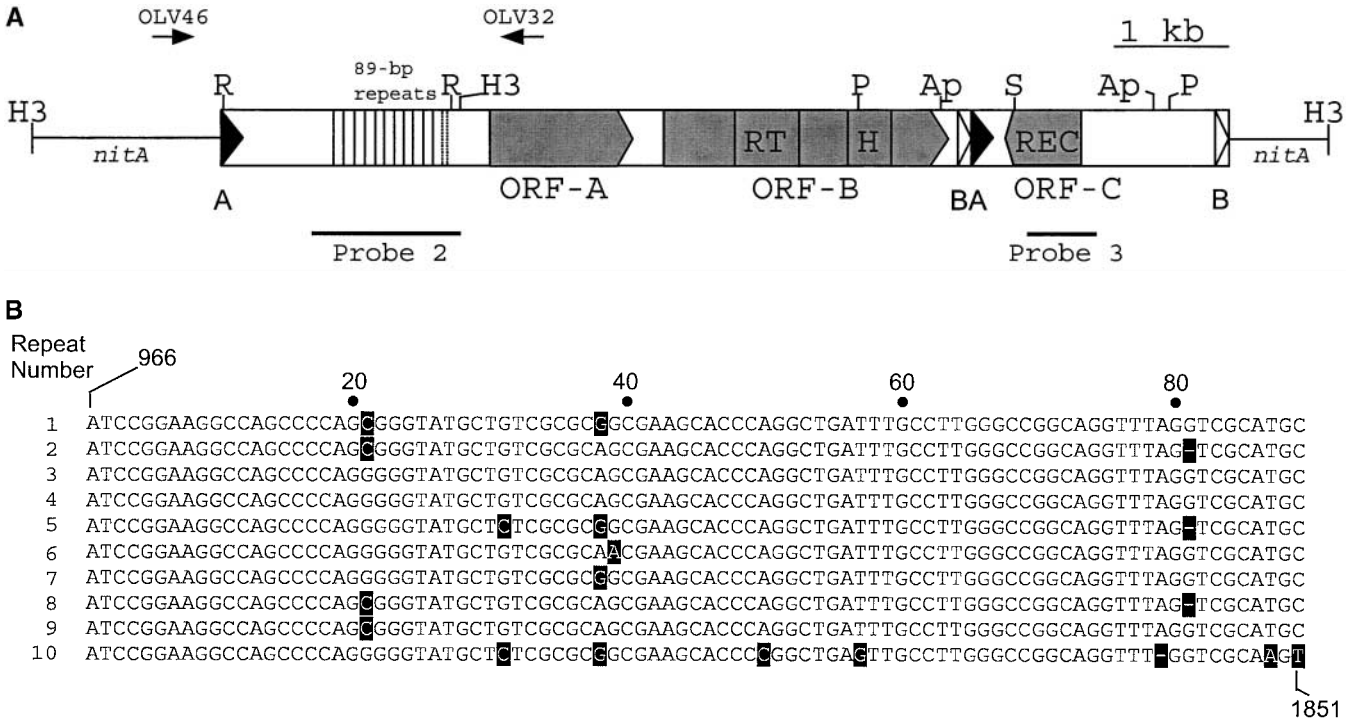

FIGURE 2.—*kangaroo-1* is an unorthodox retrotransposon containing SDR termini. (A) Diagram of *kangaroo-1* (rectangle) inserted in *nitA* (solid line). The solid and open triangles (labeled A and B) at opposite ends of the element represent the two halves of the SDR. The juxtaposed open and solid triangles (BA) represent the full-length interior direct repeat. The arrows above the diagram indicate the locations of PCR primers that were used to demonstrate that the two *Hin*dIII fragments containing *kangaroo-1* are juxtaposed as shown here. The dashed vertical lines within the 89-bp repeat region represent the last two copies of the repeat, which are less well conserved. ORFs A, B, and C are shaded. Abbreviations for RT, H, and REC are as in Figure 1. The positions of relevant restriction enzyme sites are shown: H3, *Hin*dIII; S, *Sac*I; R, *Rsa*I; P, *Pst*I; and Ap, *Apa*I. (B) Alignment of the first 10 89-bp repeats. Solid squares indicate positions that differ from the consensus. A dash indicates a gap. The numbers linked to the first and last nucleotides in the alignment correspond to the sequence of *kangaroo-1*.

ity between the RT proteins from the four DIRS1-group members extends >100 amino acids beyond the RNAse H domain (Figure 3). This conserved C-terminal extension is apparently not found in other RT proteins and may indicate that DIRS1 members share some unknown, conserved function, possibly related to their unusual mode of transposition (see below).

The *kangaroo-1* sequence includes nothing suggestive of protease or envelope functions but, like most other retrotransposons, it does contain a large ORF upstream of the RT gene (ORF-A, Figure 2A). An ORF in this location often encodes a Gag protein with one or more C_2HC "zinc-finger" motifs thought to bind nucleic acids (REIN *et al.* 1998). However, ORF-A from *kangaroo-1* lacks any discernible zinc-finger motif and BLASTP searches fail to identify any other protein with significant sequence similarity. However, our observation that ORF-A-specific transcripts are developmentally regulated (see below) suggests that ORF-A may play a role in retrotransposition. Finally, as discussed below, *kangaroo-1* also encodes a recombinase that is distinct from the integrases and endonucleases normally associated with LTR and non-LTR elements, respectively.

***kangaroo-1* is a member of a dispersed, repetitive family of mobile elements:** We compared the *kangaroo* elements of five closely related *V. carteri* f. *nagariensis* strains on DNA blots (Figure 4). Both probe 2 from the left side and probe 3 from the right side of *kangaroo-1* (see Figure 2A) recognized numerous, discrete bands in all strains (Figure 4, A and B). Many bands appeared to be present in all strains, but others were present in only one or a few of the strains examined. Most strikingly, many polymorphisms were visible between CRH7 and its clonal progenitor, EVE, which have been separated in culture for only a few years. Strains that have been isolated from one another for longer periods—such as EVE and HK9, which have been separate for at least 35 years—showed a correspondingly greater number of RFLPs. Because the restriction enzyme/probe combinations used were chosen to produce and reveal fragments with one end derived from *kangaroo-1* and the other end derived from flanking DNA, these results are consistent with *kangaroo-1* being a member of a large family of dispersed mobile elements.

Most members of the *kangaroo* family have similar structures: To determine whether other *kangaroo* ele-

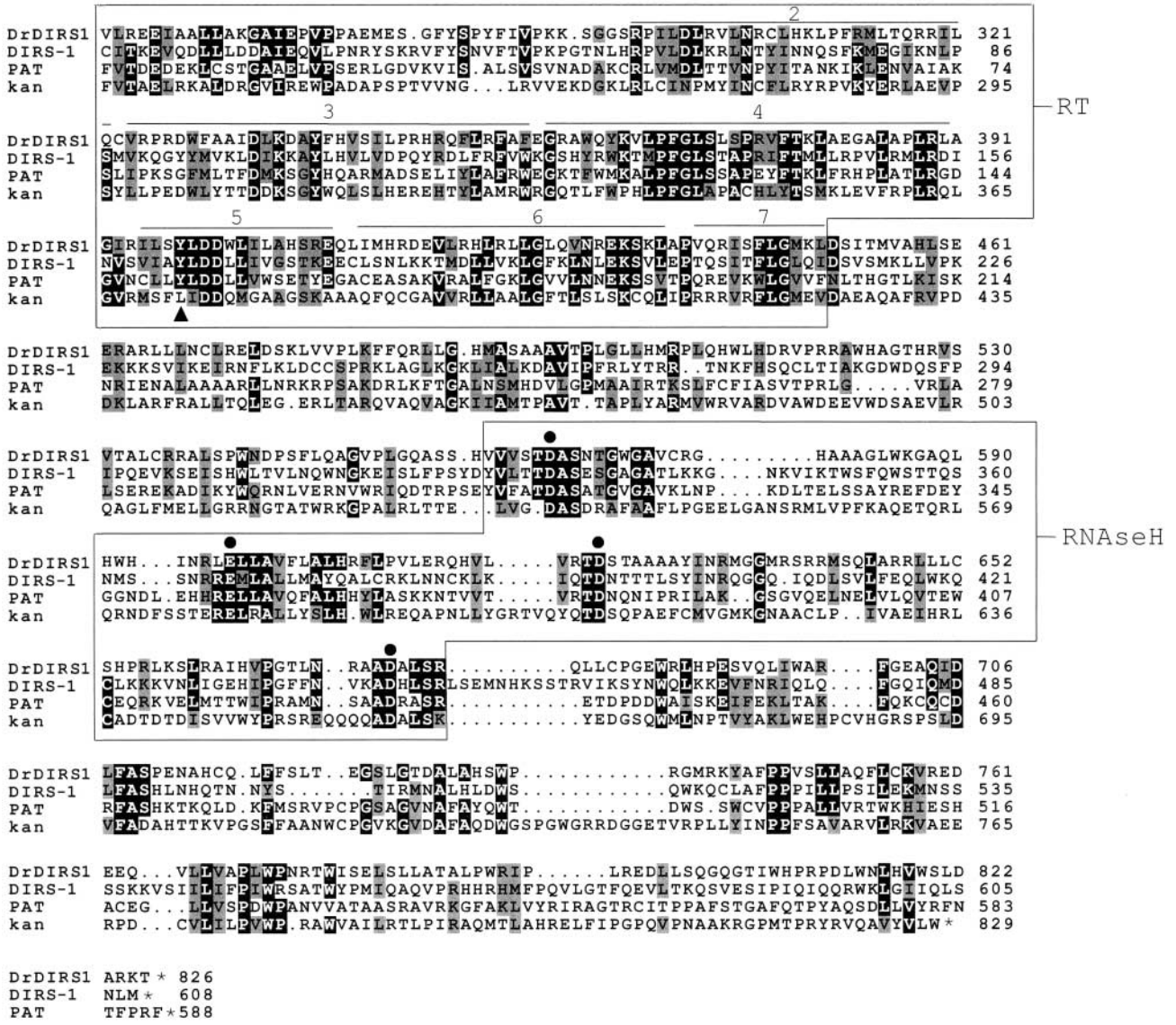

FIGURE 3.—The RT from *kangaroo-1* is closely related to the RT present in other members of the DIRS1 group. A portion of the ORF-B predicted amino acid sequence from *kangaroo-1* (kan) is aligned with the deduced RT proteins of *DIRS-1*, *DrDIRS1*, and *PAT*. The RT and RNAse H domains are outlined. Conserved regions 2–7 within the RT superfamily (XIONG and EICKBUSH 1990) are indicated by bars above the alignment. The solid triangle in region 5 indicates the first residue of the (Y/F)XDD box (see text). The alignment begins slightly upstream of region 2 and continues to the termination codon of each ORF (asterisk). The solid circles within the RNAse H domain indicate residues known to be present in the enzyme's active site and believed to be important for catalysis. The alignment was created using CLUSTALX (with manual refinement of the RNAse H domain) and MacBoxshade. Identical and similar amino acids are indicated by black and gray shading, respectively, using a 75% consensus threshold. A dot represents a gap.

ments in the genome possessed the same general structure as *kangaroo-1*, we probed a blot of *Rsa*I-digested EVE DNA with probe 2, which covers the region of *kangaroo-1* containing the 89-bp repeats (see Figure 2A). The predominant hybridizing band that was detected (Figure 5A, lane 2) was identical in size (~2 kb) to the hybridizing fragment produced by *Rsa*I digestion of cloned *kangaroo-1* (Figure 5A, lane 1), although numerous other bands of lower intensity were also seen. Similar results were obtained using other restriction enzymes and probe 3, which is derived from the opposite side of *kangaroo-1* (Figure 5B). We interpret these results to mean that a large fraction of the *kangaroo* elements within *V. carteri* possess the same general structure as *kangaroo-1*.

To determine whether other *kangaroo* elements have the same SDR termini as *kangaroo-1*, we cloned several distinct *kangaroo*-hybridizing DNA fragments that contain one or both ends of the retrotransposon and sequenced portions of these clones using oligonucleotides

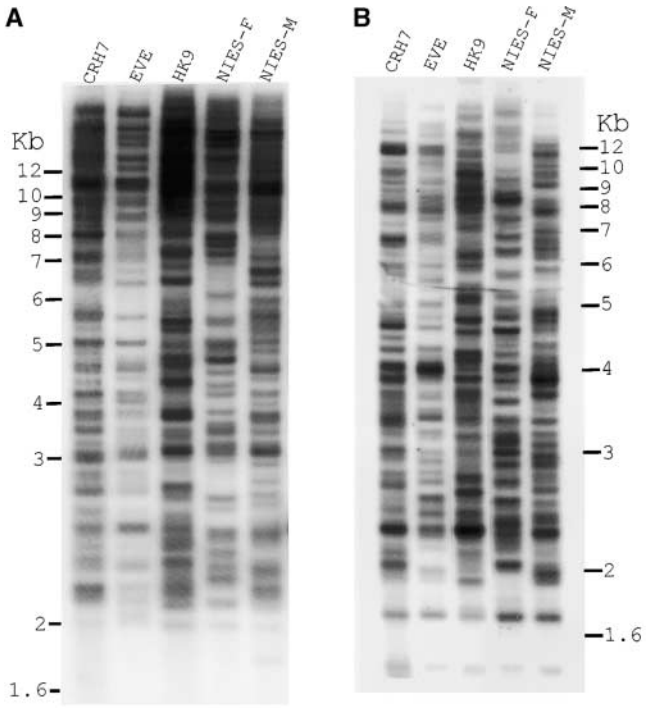

FIGURE 4.—*kangaroo-1* is a member of a dispersed repetitive family. Autoradiograms of Southern blots containing 2 μg of restricted genomic DNA from five closely related *V. carteri* f. *nagariensis* strains: CRH7, EVE, HK9, NIES female, and NIES male (see MATERIALS AND METHODS). (A) *Hin*dIII-restricted DNA hybridized with probe 2 (see Figure 2A). (B) *Sac*I-restricted DNA hybridized with probe 3 (see Figure 2A).

designed to prime just inside each terminus and read into the DNA flanking the insertion site. One of these clones, *kangaroo-2*, corresponds to a recent insertion that is present in strain LDV45, but absent in its progenitor, EVE (data not shown), providing additional evidence of *kangaroo* mobility. The other clones contained randomly selected *kangaroo*-hybridizing fragments from EVE (*kangaroo-3* and *kangaroo-6* through *kangaroo-13*) or from LDV45 (*kangaroo-4* and *kangaroo-5*) genomic libraries. Figure 5C shows an alignment of these flanking DNA:*kangaroo* junctions, with the left and right ends of *kangaroo* oriented as in Figure 2A. We found that these *kangaroo* clones are identical in sequence on the left side of the alignment beginning with the sequence 5′-TGCATGTTGATTAA-3′ and, with one exception, are also identical on the right side of the alignment until they all diverge after the sequence 5′-GACGTTTAAG CAAT-3′ (Figure 5C). We conclude that the majority of *kangaroo* elements contain SDR termini similar to those of *kangaroo-1*.

Analysis of *kangaroo* insertion sites: As shown above the alignment in Figure 5C, we observed some conservation in the DNA sequences that flank the various *kangaroo* elements that we have characterized, which suggests that *kangaroo* integration may exhibit some degree of target-site specificity. Most notably, all but one of the *kangaroo* insertions shown are bordered by the nucleotide dT on both their 5′ and 3′ ends (solid circles, Figure 5C). By comparing pre- and postintegration sites, we found that *kangaroo-1* had inserted into the sequence 5′-CTG-3′, and *kangaroo-2* had inserted into the sequence 5′-CTT-3′ (Figure 5C and data not shown). These findings lead us to conclude that the dT residue at one *kangaroo*–flanking DNA junction is derived from the target site, while the other is derived from the retrotransposon. However, there is presently no way to be certain which dT is derived from which source. Thus (as noted in the caption to Figure 5C) there is a one-nucleotide uncertainty regarding the boundaries of the retrotransposon and its target site.

***kangaroo* expression is developmentally regulated:** We isolated from *V. carteri* cDNA libraries several clones that hybridized with *kangaroo* probe 4 (Figure 6A). These clones fall into two classes. Members of the first class (clones 2, 4, 5, and 13) encode all or part of ORF-A (Figure 6A) and terminate 13–16 bp downstream of a volvocalean polyadenylation signal sequence (5′-TGTAA-3′; KIRK 1998) that is located just upstream of ORF-B. cDNA 4 contains the longest insert of this class (~2.1 kb) and corresponds to a transcript containing two exons with canonical splice sites. Interestingly, the intron includes the region of 89-bp repeats. Because cDNA 4 begins near the left end of *kangaroo-1*, it probably represents a full-length or nearly full-length cDNA clone. This suggests that a promoter may reside within the left half-repeat. The second cDNA class has only a single member, cDNA 11 (Figure 6A). This apparently partial cDNA clone corresponds to a transcript that encodes a portion of ORF-B and whose processed 3′ end terminates midway through the internal full-length repeat.

We next analyzed the accumulation of *kangaroo* transcripts by Northern blot analysis using developmentally staged RNAs harvested at various points during the asexual life cycle of *V. carteri*, which is outlined in Figure 6B. Using probe 4 (Figure 6A), we found that four major *kangaroo*-hybridizing transcripts (2.0, 3.8, 7.1, and 9.0 kb; a–d in Figure 6, C–E) are produced during development. Such transcripts were virtually undetectable in precleavage gonidia, began to accumulate during cleavage, reached a maximum level shortly after inversion, and then declined dramatically by 6 hr later (Figure 6C, lanes 1–7). These transcripts then remained at low levels throughout the rest of the asexual life cycle (time points 8–10; Figure 6E, lane 8 and data not shown). Because the transcripts reached maximum abundance during the dark period at the end of embryogenesis, but are present at much lower levels during the dark period 24 hr later (time points 9 and 10, Figure 6B; data not shown), we conclude that transcript accumulation is controlled by developmental rather than circadian factors.

Transcript a probably corresponds to cDNA 4 (Figure

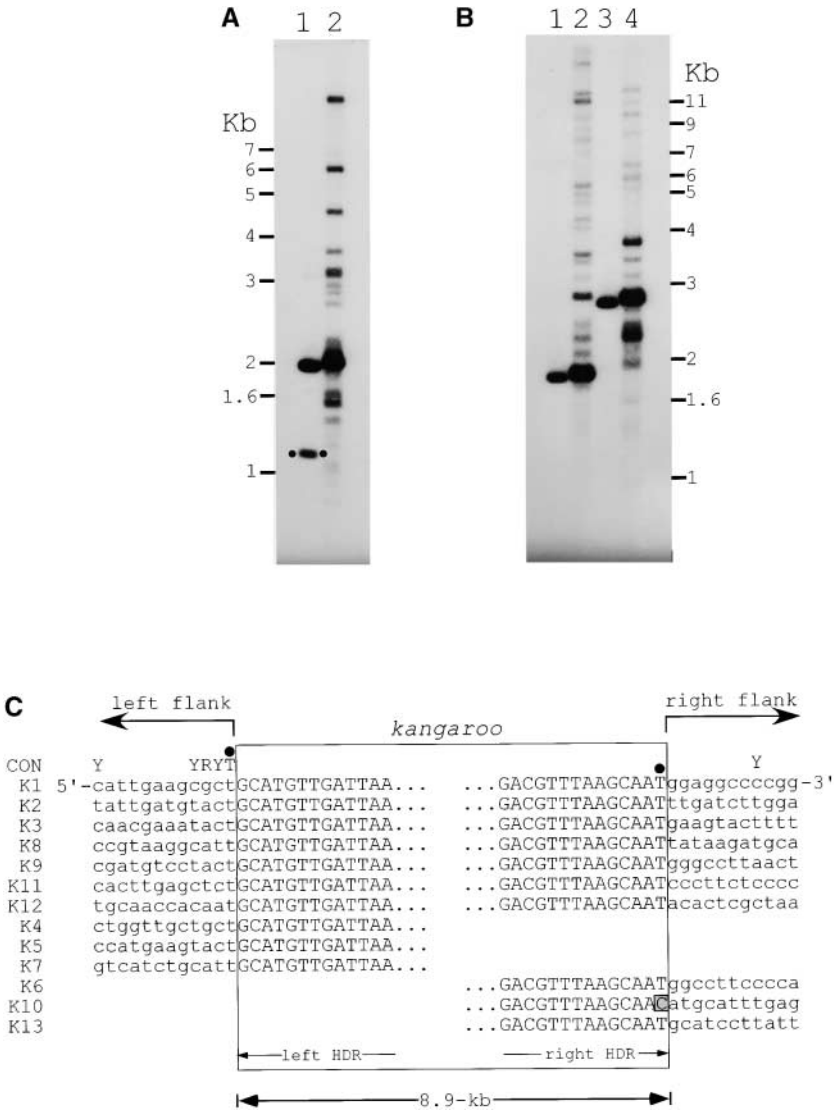

FIGURE 5.—Most *kangaroo* elements have the same general structure as *kangaroo-1*. (A) Autoradiogram of a Southern blot hybridized with probe 2 (see Figure 2A). Lane 1, 6 ng of *Rsa*I-digested pLD41 (which contains the left half of *kangaroo-1*); lane 2, 1.5 μg of EVE genomic DNA digested with *Rsa*I. The 1.2-kb hybridizing band flanked by dots in lane 1 is a fragment of pLD41 that contains only a short segment of *kangaroo-1* and whose size is unrelated to the true structure of the retrotransposon. (B) Autoradiogram of a Southern blot hybridized with probe 3 (see Figure 2A). Lane 1, 6 ng of *Apa*I-digested pLD40 (which contains the right half of *kangaroo-1*); lane 2, 1.5 μg of *Apa*I-digested EVE genomic DNA; lane 3, 6 ng of pLD40 digested with *Pst*I; lane 4, 1.5 μg of *Pst*I-digested EVE genomic DNA. (C) Aligned sequences of the retrotransposon:flanking DNA junctions of several independent *kangaroo* clones (*kangaroo-1* through *kangaroo-13*). The *kangaroo* sequences are enclosed by a rectangle and shown in uppercase. Only a short segment of each terminus of the *kangaroo* elements is shown. Flanking DNA is in lowercase and delineated by arrows. Note that (with one exception) the *kangaroo* elements are flanked by a dT residue on both sides (solid circles), and it is unclear which residue is derived from the retroelement and which is derived from the target DNA; possibly, the rectangle delineating the boundaries of *kangaroo* should be shifted 1 bp to the left. HDR, half direct repeat; CON, conserved target DNA sequences; Y, pyrimidine; R, purine.

6A), because it is of the appropriate size (~2.1 kb), and it hybridizes with ORF-A-specific probe 6 (data not shown) but not with intron-specific probe 7 (Figure 6E, lane 6), or with ORF-B-specific probe 5 (Figure 6D, lane 6), or with probe 3, which is derived from the right side of *kangaroo* (data not shown).

Transcript d is about the right size to be a full-length, unspliced *kangaroo* RNA species (Figure 6A). Consistent with this view, transcript d hybridizes with all of the *kangaroo* probes that we have used in Northern blotting experiments (Figure 6, C–E and data not shown). Presumably, transcript d is the template used during reverse transcription.

The discovery of an intron within *kangaroo-1* prompted us to also examine the nature of transcripts b and c. We found that the hybridization patterns for transcript b are identical to those of transcript a (Figure 6, C and D and data not shown) except that b also hybridizes with the intron-specific probe 7 (Figure 6E, lane 6); we also found that the hybridization patterns for transcript c are identical to those of transcript d (Figure 6, C and D, and data not shown) except that c fails to hybridize to the intron-specific probe 7 (Figure 6E, lane 6). These results lead us to propose that transcripts b and c have the structures shown in Figure 6A. The sizes of transcripts b and c are consistent with this interpretation.

Finally, because cDNA 11 (Figure 6A) apparently does not correspond to any of the major *kangaroo*-hybridizing transcripts visualized by Northern blot analysis, it seems likely that it represents a relatively low-abundance message.

***kangaroo* may integrate as a closed-circle, double-stranded DNA copy:** A common feature of LTR and non-LTR retrotransposon integration is the generation of element-specific target-site duplications (GABRIEL and BOEKE 1993). Thus, one interpretation of the finding that *kangaroo* elements are bounded by single dT residues (Figure 5C) is that *kangaroo* may integrate using a conventional retrotransposon mechanism that duplicates the conserved dT target-site residue. However, to

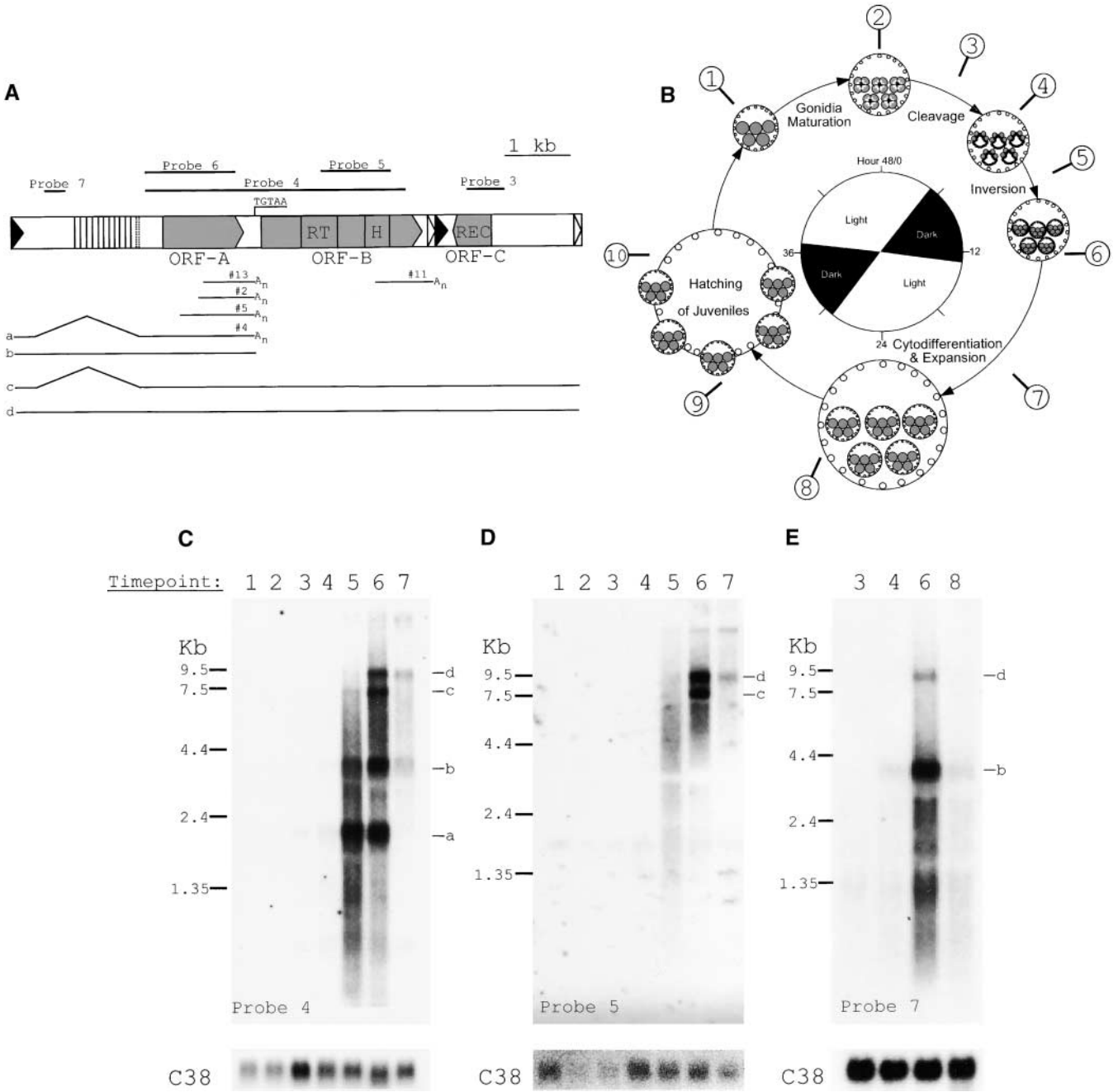

FIGURE 6.—Multiple developmentally regulated transcripts from *kangaroo* are observed. (A) A diagram of *kangaroo-1* showing the extent of five *kangaroo* cDNA clones (2, 4, 5, 11, and 13) and the locations of sequences used as probes for the Northern blots shown in C–E. Also shown are the deduced structures of the four major *kangaroo* transcripts (a–d) shown in C–E. TGTAA, volvocalean polyadenylation signal (KIRK 1998); A_n, poly(A) tail. (B) The asexual life cycle of *V. carteri* (KIRK 1998). An individual *V. carteri* spheroid (time point 1) contains only two cell types: large asexual reproductive cells called gonidia in the interior of the sphere and small, terminally differentiated somatic cells at its surface. A 24-hr light-dark cycle (inner circle) can be used to synchronize development. Under these conditions, gonidia become mature and begin to divide near the end of one light period, and embryogenesis is completed in the dark. Following the completion of mitotic divisions, the embryo turns inside out in a process called inversion to produce a juvenile, which is a miniature version of an adult spheroid. During the second light period, the two cell types of the juvenile differentiate and both the juveniles and parental spheroids expand by deposition of extracellular matrix (ECM). Near the end of the second dark period, the juveniles digest holes in the parental ECM and swim away. The cycle is completed when the gonidia of the juvenile initiate a new round of embryogenesis. The 10 time points at which total RNA was isolated from synchronized gonidia, embryos, or juveniles are indicated and correspond to the lane numbers shown in C–E. Somatic cells from the parental spheroids were removed from all samples prior to RNA isolation. (C and D) Autoradiograms of Northern blots containing 10 μg of total RNA isolated at time points 1–7 as indicated in B and hybridized with the indicated probes. (E) Autoradiogram of a Northern blot containing 8 μg of total RNA isolated at time points 3, 4, 6, and 8 and hybridized with probe 7. As an RNA-loading control, the same blots were stripped and rehybridized with a probe for a transcript of constant abundance, C38 (bottom).

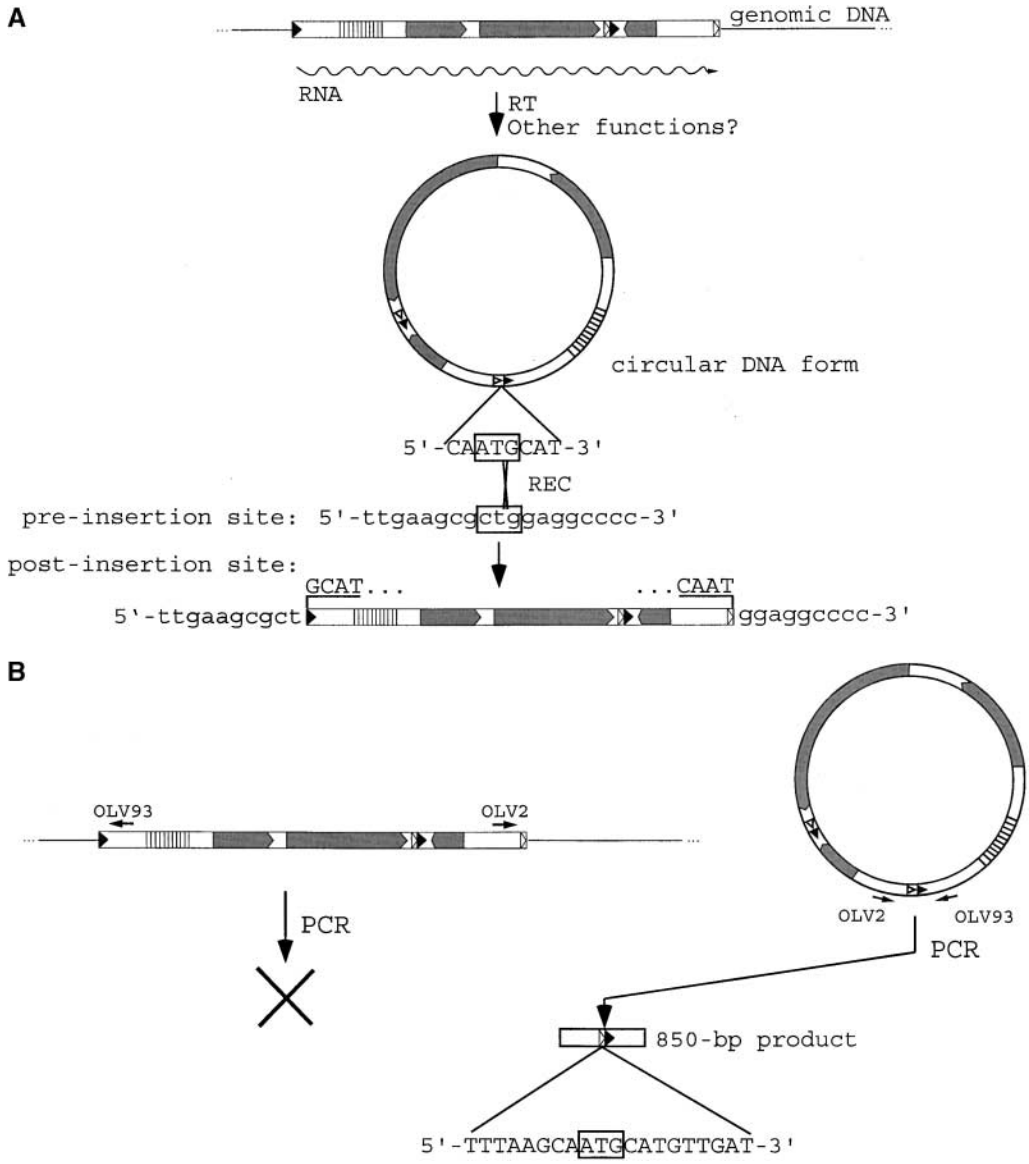

FIGURE 7.—A model for *kangaroo* insertion. (A) Comparison of the observed pre- and postinsertion target-site sequences for *kangaroo-1*. The postinsertion product could be generated if integration occurred by a single DNA crossover event between the target site (boxed 5′-ctg-3′) and the circle junction that is created by ligation of the ends of the linear form of *kangaroo* (boxed 5′-ATG-3′). Here, the crossover (black X) is shown as being 3′ to the dT residues present in the recombining DNAs. Alternatively, the same postinsertion product could be generated if the DNA crossover occurred 5′ to the dT residues (not shown). This recombination event could be catalyzed by the *kangaroo* Rec protein (see text). (B) PCR amplification of the *kangaroo* circle junction region. PCR amplification of linear, integrated forms of *kangaroo* with OLV93 and OLV2 would not be expected to produce a product (left), but amplification from a closed-circle DNA form of *kangaroo* would produce a product with a precisely defined size and junction (boxed 5′-ATG-3′) between the two newly juxtaposed half-repeats (right). A PCR product whose size and sequence is fully consistent with the existence of such a circular form was obtained (see text).

our knowledge, no LTR or non-LTR retrotransposon is known to create a single-base-pair target-site duplication. More importantly, *kangaroo* and other members of the DIRS1 group do not encode a conventional DDE-integrase or a non-LTR-like endonuclease. These results and the asymmetric nature of *kangaroo*'s termini lead us to propose the insertional model presented in Figure 7A, which involves an extrachromosomal DNA circle very similar to the one that was proposed as an intermediate in the replication of *DIRS-1* (CAPPELLO *et al.* 1985) and a second, related element called *TOC1* (DAY *et al.* 1988; see below). Specifically, it is postulated that a critical intermediate formed during retrotransposition of *kangaroo* is an extrachromosomal, closed-circle, double-stranded DNA copy in which the ends of the retroelement are fused to generate an identical second copy of the full, interior direct repeat (Figure 7A). As shown in Figure 7A, the integration product that is observed for *kangaroo-1* (Figure 5C) could be produced by a DNA crossover event between the newly formed circle junction of *kangaroo* and the target site. The integration product produced by insertion of *kangaroo-2* (Figure 5C) can be explained in a similar manner (not shown).

In accord with this model, we were able to use the PCR primers shown in Figure 7B to amplify a fragment from EVE DNA whose size and sequence is fully consistent with the product that would be produced from the postulated closed-circle DNA intermediate—but not from a linear—DNA form of *kangaroo*. It is formally possible that such a product could have been produced by amplification from two copies of *kangaroo* integrated in the genome in a head-to-tail manner. However, we believe that even if such a tandem juxtaposition of linear *kangaroo* elements were to exist in the genome, it would

be unlikely to have coincidentally generated a sequence identical to that which would be created by ligation of the half-repeats in a circular *kangaroo* form.

***kangaroo* encodes a protein related to λ-site-specific integrase:** Goodwin and Poulter (2001) reported that members of the DIRS1 group of retrotransposons potentially encode a protein (ORF-C, Figure 1B) related to the Int family of recombinases (Nunes-Duby *et al.* 1998). The quintessential member of this protein family is the integrase protein from bacteriophage λ, which catalyzes the integration and excision of the λ-genome at a specific site within the *Escherichia coli* chromosome (Landy 1989). Int family members also catalyze the non-site-specific integration of circular forms of conjugative (DNA) transposons (Scott and Churchward 1995). It is important to note that the Int recombinases are distinct in structure and catalytic mechanism from the DDE-type integrases normally associated with retroviruses and LTR retrotransposons (Haren *et al.* 1999) and from the endonucleases associated with non-LTR elements (Furano 2000). Members of the Int family are quite divergent in sequence. Indeed, only four amino acids, which comprise the "RHRY" tetrad essential for catalysis, are present in all members of this group (Nunes-Duby *et al.* 1998). In support of the results of Goodwin and Poulter (2001) we have found that *kangaroo* also potentially encodes a 225-amino-acid protein related to λ-integrase (ORF-C, Figure 2A). Figure 8 shows part of the deduced ORF-C protein from *kangaroo-1* aligned with portions of λ-integrase and several other members of the Int family, including the deduced recombinase proteins from *PAT*, *DIRS-1*, and *DrDIRS1*. The alignment identifies the conserved box I and box II regions containing the catalytic RHRY tetrad (Nunes-Duby *et al.* 1998) and, in extension of the results of Goodwin and Poulter (2001), we have found that the recombinase proteins from the DIRS1 family also contain the "Patch II" and "Patch III" regions identified by Nunes-Duby *et al.* (1998; Figure 8). Although the ORF-C protein from *kangaroo-1* is predicted to contain several insertions and deletions of amino acids relative to other proteins shown in Figure 8 (*e.g.*, immediately following box I), these insertions/deletions are likely to fall within loop regions of the folded protein (Nunes-Duby *et al.* 1998) and presumably do not prevent its adopting a structure similar to λ-integrase. These results lead us to conclude that ORF-C encodes a member of the Int family and to speculate that this enzyme catalyzes the integration of *kangaroo* via the mechanism we have proposed in Figure 7A.

DISCUSSION

kangaroo is a member of a newly recognized but rapidly growing class of retrotransposons—the DIRS1 group—whose members share structures, genes, and (probably) integration mechanisms that differ from those of the LTR and the conventional non-LTR retroelements. The first members of this group to be recognized were *DIRS-1*, from Dictyostelium (Cappello *et al.* 1985), and *PAT*, from *P. redivivus* (de Chastonay *et al.* 1992), which were united solely by the similarity of their RT proteins and appeared to differ greatly in the structure of their termini. A less-well-characterized, but clearly related, retroelement called *Prt1* was also found in the fungus *Phycomyces blakesleeanus* (Ruiz-Perez *et al.* 1996). Here we have characterized the first member of the group to be identified in a photoautotroph, namely *kangaroo-1* of *V. carteri*, a green alga, and we have shown that *kangaroo* not only is closely related to *DIRS-1* and *PAT* in terms of RT amino acid sequence, but also has split direct repeat termini very similar in structure to the termini of *PAT*. Meanwhile, Goodwin and Poulter (2001) have reported discovering several additional metazoan members of the DIRS1 group by screening public databases. These included *DrDIRS1* of the zebrafish *Danio rerio*, *TnDIRS1* of the pufferfish *Tetraodon nigroviridis*, and *CbPAT1* of *Caenorhabditis briggsae*, as well as several fragmentary *DIRS1*-like sequences from two species of Xenopus and the sea urchin, *Strongylocentrotus purpuratus* (Goodwin and Poulter 2001; L. Duncan, unpublished data). DIRS1-group retroelements have not yet been found in insects or vascular plants.

However, earlier studies (Day *et al.* 1988; Day and Rochaix 1991) had revealed that *Chlamydomonas reinhardtii*, the closest unicellular relative of *V. carteri*, contains a mobile element called *TOC1* with SDRs similar in structure (but not sequence) to those of *kangaroo-1* and *PAT* (Figure 1B). *TOC1* also contains a stretch of 76-bp repeats that is strikingly similar in position and length to the stretch of 89-bp repeats in *kangaroo-1* (compare Figures 1B and 2A), although the nucleotide sequences of their repetitive units are wholly unrelated. Because *TOC1* does not encode RT or other proteins that play a role in retrotransposition, it appears to be a nonautonomous element and could not be included in the phylogenetic analysis of retroelement RT sequences. Nevertheless, the studies reported here lead us to predict that *TOC1* probably replicates and transposes by a mechanism similar to the one that we have proposed for *kangaroo* in Volvox, using functions that are encoded by an autonomous *kangaroo*-like element elsewhere in the Chlamydomonas genome. Consistent with this hypothesis, we have found that several Chlamydomonas expressed sequence tags are present in the public databases that encode peptides with significant similarity to regions of the *kangaroo* RT protein, including the C-terminal extension that we believe is diagnostic for the DIRS1 family of transposons (L. Duncan, unpublished data). This, in turn, leads us to suspect that *kangaroo*-like elements may be widely distributed within the order Volvocales. Indeed, with representatives now known to be present in slime molds, fungi, green algae, and a variety of different metazoans, it

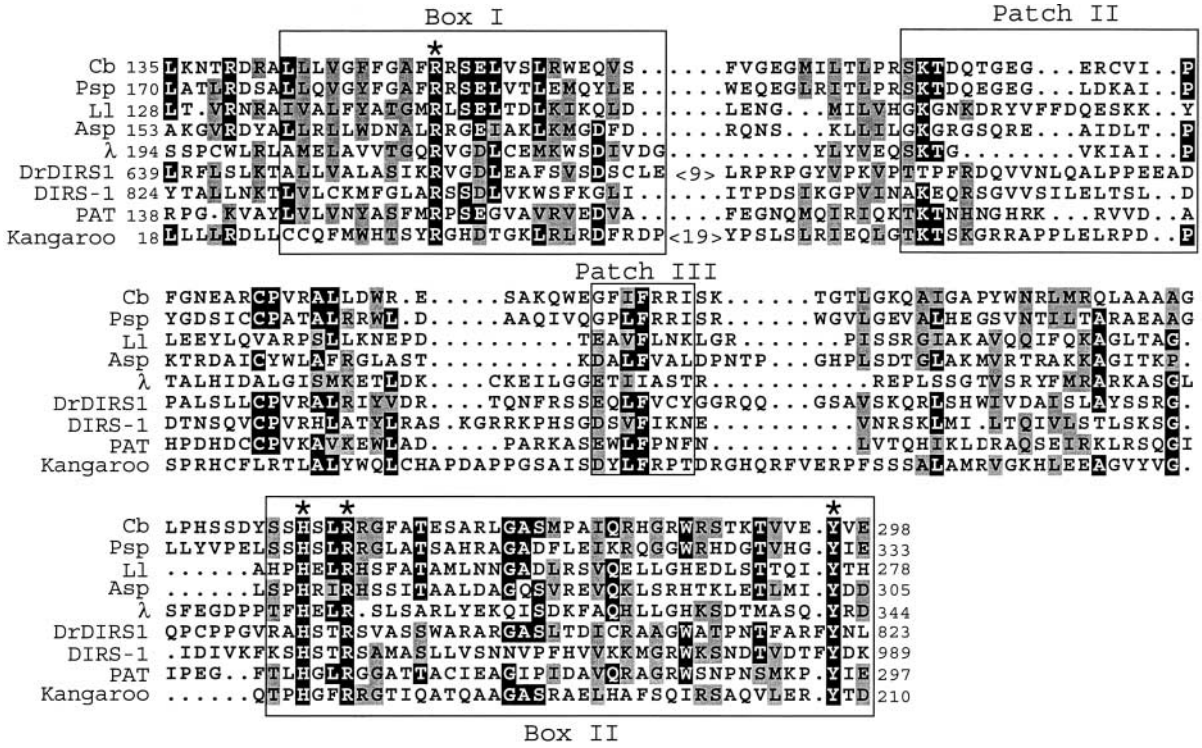

FIGURE 8.—DIRS1-group members encode a recombinase related to λ-site-specific integrase. The predicted ORF-C proteins from *kangaroo-1*, *DIRS-1*, *DrDIRS1*, and *PAT* are aligned with a number of recombinase proteins from the Int family. The conserved residues forming the RHRY catalytic tetrad are indicated by asterisks. For clarity, the sequences FGPDYSHVI and RGGGPFRGFPLPLPDPFGA were removed from the *DrDIRS1* and *kangaroo* proteins, respectively, at the positions indicated just to the right of box I. The other aligned proteins are: Cb (*Coxiella burnetii* integrase, CAA75853), Psp (Pseudomonas sp. integrase, CAA67462), Ll (*Lactobacillus leichmannii* XerC, CAA59018), Asp (Anabaena sp. integrase, BAB77331), and λ (integrase, P03700). The alignment was created using CLUSTALX (with manual refinement) and MacBoxshade. Identical and similar residues are shaded black and gray, respectively, using a 50% consensus threshold. A dot indicates a gap.

seems reasonable to postulate that DIRS1-group elements may be almost as widely distributed among the eukaryotes as the LTR and conventional non-LTR retroelements are now known to be.

The two DIRS1-group subfamilies contain related terminal structures: As discussed in the Introduction, most RT-based phylogenies identify clades of retroelements that share other important structural and genetic features. Thus, we were initially surprised to find that the DIRS1 clade grouped elements that apparently contain dissimilar termini: namely, those containing inverted repeats and those containing SDRs. However, we subsequently realized that the elements with inverted-repeat termini have other features that are structurally similar to the SDRs found in *kangaroo*, *PAT*, and *TOC1*. Specifically, both *DIRS-1* and *DrDIRS1* have short sequences (labeled A and B in Figure 1B) that extend past the end of, or are part of, the terminal inverted repeats, and these same sequences are also found juxtaposed internally in an inverted orientation (A′B′) in a structure called the internal complementary region (ICR; CAPPELLO *et al.* 1985). *TnDIRS1* also has a similar structure (GOODWIN and POULTER 2001). This arrangement can be symbolized as: A > . . . < A < B . . . B >, and we propose that it may represent an altered form of the type of SDR seen in *PAT*, *TOC1*, and *kangaroo*, which can be symbolized as: A > . . . B > A > . . . B >. It is tempting to speculate that one of the DIRS1-group subfamilies was derived from the other subfamily by inversion of the full interior repeat (Figure 1B).

DIRS1 elements appear to integrate by a novel mechanism: Here we have proposed that members of the DIRS1 group may transpose by a mechanism similar to the one proposed initially by CAPPELLO *et al.* (1985) and DAY *et al.* (1988), in which RT generates a closed-circle, double-stranded DNA intermediate that is then inserted into a target site by a single crossover event. If this model is correct, then the integration mechanism used by the DIRS1-group members would be clearly distinct from that used during insertion of known LTR and conventional non-LTR retrotransposons.

This model was initially based on: (1) our observation that members of the DIRS1 group contain unorthodox, but related, terminal structures and do not encode a conventional retrotransposon integrase or endonuclease; (2) our comparisons of *kangaroo* pre- and postinte-

gration sites; and (3) our demonstration that we could amplify a PCR product from Volvox DNA whose structure is consistent with the existence of the postulated closed-circular DNA intermediate. These latter two results constitute the first pieces of experimental evidence in support of this kind of model.

Meanwhile, Goodwin and Poulter (2001) proposed a very similar model, after discovering that members of the DIRS1 group encode proteins of the Int family of recombinases, some of which (such as λ-site-specific integrase) are known to mediate this type of integration process. We have now shown that *kangaroo* also encodes such a recombinase. Thus, our study and that of Goodwin and Poulter (2001) are mutually reinforcing.

However, whereas Goodwin and Poulter (2001) consider the DIRS1 elements as a "group of LTR retrotransposons," we believe that because of their very substantial differences from LTR elements in termini, gene content, and probable integration mechanisms, the DIRS1 group should be considered to be a third class of retrotransposons, distinct from both the LTR and the traditional non-LTR classes.

Developmental regulation of *kangaroo* expression: We have shown that the accumulation of four discrete transcripts produced by *kangaroo* is developmentally regulated, which is a property shared with numerous other retrotransposons. The expression of at least 19 different Drosophila retroelements is controlled both temporally and spatially during development (Ding and Lipshitz 1994; Frommer *et al.* 1994; Mozer and Benzer 1994; Bronner *et al.* 1995; Awasaki *et al.* 1996; Kerber *et al.* 1996; Marsano *et al.* 2000), as is expression of the zebrafish retroelement *bhikhara* (Vogel and Gerster 1999) and the Xenopus *1A11* element (Greene *et al.* 1993). Similarly, transcription of *Ty* elements in *Saccharomyces cerevisiae* (Errede *et al.* 1987), LINE-1 in mammals (Ostertag and Kazazian 2001), and *DIRS-1* in *D. discoideum* (Cohen *et al.* 1984) all appear to be under cell-type or stage-specific controls. As with expression of more conventional kinds of genes, developmentally regulated expression of retrotransposons has been shown to involve both *cis*-regulatory elements located within the transposons themselves and *trans*-acting factors encoded elsewhere (Errede *et al.* 1987; Ding and Lipshitz 1994; Mozer and Benzer 1994; Bronner *et al.* 1995; Awasaki *et al.* 1996; Kerber *et al.* 1996; Vogel and Gerster 1999). It has been argued that the developmentally regulated expression of such retroelements may have been selected for by virtue of its benefit to the host (Ding and Lipshitz 1994; Bronner *et al.* 1995), but convincing evidence in favor of this hypothesis remains lacking.

Interestingly, observations made in the related unicell *C. reinhardtii* raise the possibility that developmental accumulation of *kangaroo* transcripts could be regulated epigenetically. Jeong *et al.* (2002) have shown that *TOC1* expression in *C. reinhardtii* is repressed by two genes (*Mut9* and *Mut11*) that are involved in transcriptional gene silencing (TGS). Furthermore, the levels of *TOC1* RNA are also controlled, at least in part, by degradation that is dependent on the *Mut6* RNA helicase (Wu-Scharf *et al.* 2000), which is a component of the post-transcriptional gene silencing (PTGS) machinery of *C. reinhardtii*. It is conceivable that *kangaroo* expression in *V. carteri* is controlled in an analogous manner by uncharacterized, developmentally regulated TGS or PTGS mechanisms.

***kangaroo* as a molecular genetic tool:** At present, the only method available for cloning genes by forward genetics in *V. carteri* has involved tagging with the DNA transposon, *Jordan* (Miller *et al.* 1993). Although this approach has been used successfully to clone several developmentally important genes (Kirk *et al.* 1999; Miller and Kirk 1999; I. Nishii, personal communication), we have encountered several cases in which interesting mutations that have the earmarks of transposon-induced mutations cannot be correlated with *Jordan* RFLPs (our unpublished observations). Evidence presented here identifies *kangaroo* as a second, highly mobile element within the *V. carteri* genome that is capable of causing gene disruptions. Whether *kangaroo* will turn out to have other properties required to make it a second useful transposon-tagging tool for *V. carteri* developmental biologists, only time will tell.

We thank I. Nishii for staged RNAs and the Northern blot (with associated C38 loading control data) used in Figure 6C; A. Horton and D. Berg for helpful advice; and J. Umen, S. Miller, I. Nishii, and C. Shaffer for comments on the manuscript. L.D. was a postdoctoral fellow of, and this investigation has been aided by, a grant from the Jane Coffin Childs Memorial Fund for Medical Research. This work was also supported by a National Science Foundation grant (IBN-9904739) to D.K.

LITERATURE CITED

Awasaki, T., N. Juni and K. M. Yoshida, 1996 An eye imaginal disc-specific transcriptional enhancer in the long terminal repeat of the *tom* retrotransposon is responsible for eye morphology mutations of *Drosophila ananassae*. Mol. Gen. Genet. **251:** 161–166.

Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller *et al.*, 2002 The Pfam protein families database. Nucleic Acids Res. **30:** 276–280.

Bronner, G., H. Taubert and H. Jackle, 1995 Mesoderm-specific *B104* expression in the Drosophila embryo is mediated by internal *cis*-acting elements of the transposon. Chromosoma **103:** 669–675.

Cappello, J., K. Handelsman and H. F. Lodish, 1985 Sequence of Dictyostelium *DIRS-1*: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. Cell **43:** 105–115.

Coffin, J., 1993 Reverse transcriptase and evolution, pp. 445–479 in *Reverse Transcriptase*, edited by A. M. Skalka and S. P. Goff. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cohen, S. M., J. Cappello and H. F. Lodish, 1984 Transcription of *Dictyostelium discoideum* transposable element *DIRS-1*. Mol. Cell. Biol. **4:** 2332–2340.

Craig, N. L., 1997 Target site selection in transposition. Annu. Rev. Biochem. **66:** 437–474.

Day, A., and J. D. Rochaix, 1991 A transposon with an unusual LTR arrangement from *Chlamydomonas reinhardtii* contains an internal tandem array of 76 bp repeats. Nucleic Acids Res. **19:** 1259–1266.

Day, A., M. Schirmer-Rahire, M. R. Kuchka, S. P. Mayfield and

J. D. Rochaix, 1988 A transposon with an unusual arrangement of long terminal repeats in the green alga *Chlamydomonas reinhardtii*. EMBO J. **7:** 1917–1927.

de Chastonay, Y., H. Felder, C. Link, P. Aeby, H. Tobler *et al.*, 1992 Unusual features of the retroid element *PAT* from the nematode *Panagrellus redivivus*. Nucleic Acids Res. **20:** 1623–1628.

Ding, D., and H. D. Lipshitz, 1994 Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis. Genet. Res. **64:** 167–181.

Doolittle, R. F., D. F. Feng, M. S. Johnson and M. A. McClure, 1989 Origins and evolutionary relationships of retroviruses. Q. Rev. Biol. **64:** 1–30.

Eickbush, T. H., 1997 Telomerase and retrotransposons: Which came first? Science **277:** 911–912.

Errede, B., M. Company and C. A. Hutchison, 1987 *Ty1* sequence with enhancer and mating-type-dependent regulatory activities. Mol. Cell. Biol. **7:** 258–265.

Finnegan, D. J., 1989 Eukaryotic transposable elements and genome evolution. Trends Genet. **5:** 103–107.

Frommer, G., R. Schuh and H. Jackle, 1994 Localized expression of a novel *micropia*-like element in the blastoderm of *Drosophila melanogaster* is dependent on the anterior morphogen bicoid. Chromosoma **103:** 82–89.

Furano, A. V., 2000 The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog. Nucleic Acid Res. Mol. Biol. **64:** 255–294.

Gabriel, A., and J. D. Boeke, 1993 Retrotransposon reverse transcription, pp. 275–328 in *Reverse Transcriptase*, edited by A. M. Skalka and S. P. Goff. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Goodwin, T. J., and R. T. Poulter, 2001 The DIRS1 group of retrotransposons. Mol. Biol. Evol. **18:** 2067–2082.

Greene, J. M., H. Otani, P. J. Good and I. B. Dawid, 1993 A novel family of retrotransposon-like elements in *Xenopus laevis* with a transcript inducible by two growth factors. Nucleic Acids Res. **21:** 2375–2381.

Gruber, H., S. D. Goetinck, D. L. Kirk and R. Schmitt, 1992 The nitrate reductase-encoding gene of *Volvox carteri*: map location, sequence and induction kinetics. Gene **120:** 75–83.

Haren, L., N. Ton-Hoang and M. Chandler, 1999 Integrating DNA: transposases and retroviral integrases. Annu. Rev. Microbiol. **53:** 245–281.

Harper, J. F., K. S. Huson and D. L. Kirk, 1987 Use of repetitive sequences to identify DNA polymorphisms linked to *regA*, a developmentally important locus in Volvox. Genes Dev. **1:** 573–584.

Jeong, B., D. Wu-Scharf, C. Zhang and H. Cerutti, 2002 Suppressors of transcriptional transgenic silencing in Chlamydomonas are sensitive to DNA-damaging agents and reactivate transposable elements. Proc. Natl. Acad. Sci. USA **99:** 1076–1081.

Kerber, B., S. Fellert, H. Taubert and M. Hoch, 1996 Germ line and embryonic expression of *Fex*, a member of the Drosophila *F*-element retrotransposon family, is mediated by an internal *cis*-regulatory control region. Mol. Cell. Biol. **16:** 2998–3007.

Kirk, D., 1998 *Volvox: Molecular-Genetic Origins of Multicellularity and Cellular Differentiation*. Cambridge University Press, Cambridge, UK.

Kirk, D. L., and M. M. Kirk, 1983 Protein synthetic patterns during the asexual life cycle of *Volvox carteri*. Dev. Biol. **96:** 493–506.

Kirk, M. M., and D. L. Kirk, 1985 Translational regulation of protein synthesis, in response to light, at a critical stage of Volvox development. Cell **41:** 419–428.

Kirk, M. M., K. Stark, S. M. Miller, W. Muller, B. E. Taillon *et al.*, 1999 *regA*, a Volvox gene that plays a central role in germsoma differentiation, encodes a novel regulatory protein. Development **126:** 639–647.

Kohlstaedt, L. A., J. Wang, P. A. Rice, J. M. Friedman and T. A. Steitz, 1993 The structure of HIV-1 reverse transcriptase, pp. 223–249 in *Reverse Transcriptase*, edited by A. M. Skalka and S. P. Goff. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kumar, A., and J. L. Bennetzen, 1999 Plant retrotransposons. Annu. Rev. Genet. **33:** 479–532.

Landy, A., 1989 Dynamic, structural, and regulatory aspects of λ site-specific recombination. Annu. Rev. Biochem. **58:** 913–949.

Lindauer, A., D. Fraser, M. Bruderlein and R. Schmitt, 1993 Reverse transcriptase families and a *copia*-like retrotransposon, *Osser*, in the green alga *Volvox carteri*. FEBS Lett. **319:** 261–266.

Malik, H. S., and T. H. Eickbush, 2001 Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses. Genome Res. **11:** 1187–1197.

Marsano, R. M., R. Moschetti, C. Caggese, C. Lanave, P. Barsanti *et al.*, 2000 The complete *Tirant* transposable element in *Drosophila melanogaster* shows a structural relationship with retrovirus-like retrotransposons. Gene **247:** 87–95.

McClure, M. A., 1993 Evolutionary history of reverse transcriptase, pp. 425–444 in *Reverse Transcriptase*, edited by A. M. Skalka and S. P. Goff. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Miller, S. M., and D. L. Kirk, 1999 *glsA*, a Volvox gene required for asymmetric division and germ cell specification, encodes a chaperone-like protein. Development **126:** 649–658.

Miller, S. M., R. Schmitt and D. L. Kirk, 1993 *Jordan*, an active Volvox transposable element similar to higher plant transposons. Plant Cell **5:** 1125–1138.

Mozer, B. A., and S. Benzer, 1994 Ingrowth by photoreceptor axons induces transcription of a retrotransposon in the developing Drosophila brain. Development **120:** 1049–1058.

Nunes-Duby, S. E., H. J. Kwon, R. S. Tirumalai, T. Ellenberger and A. Landy, 1998 Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic Acids Res. **26:** 391–406.

Ostertag, E. M., and H. H. Kazazian, Jr., 2001 Biology of mammalian *L1* retrotransposons. Annu. Rev. Genet. **35:** 501–538.

Rein, A., L. E. Henderson and J. G. Levin, 1998 Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication. Trends Biochem. Sci. **23:** 297–301.

Ruiz-Perez, V. L., F. J. Murillo and S. Torres-Martinez, 1996 *Prt1*, an unusual retrotransposon-like sequence in the fungus *Phycomyces blakesleeanus*. Mol. Gen. Genet. **253:** 324–333.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SanMiguel, P., A. Tikhonov, Y. K. Jin, N. Motchoulskaia, D. Zakharov *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

Scott, J. R., and G. G. Churchward, 1995 Conjugative transposition. Annu. Rev. Microbiol. **49:** 367–397.

Smit, A. F., 1999 Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr. Opin. Genet. Dev. **9:** 657–663.

Starr, R. C., 1969 Structure, reproduction and differentiation in *Volvox carteri* f. *nagariensis* Iyengar, strains HK9 and HK10. Arch. Protistenkd. **111:** 204–222.

Tam, L. W., and D. L. Kirk, 1991 Identification of cell-type-specific genes of *Volvox carteri* and characterization of their expression during the asexual life cycle. Dev. Biol. **145:** 51–66.

Vogel, A. M., and T. Gerster, 1999 Promoter activity of the zebrafish *bhikhari* retroelement requires an intact activin signaling pathway. Mech. Dev. **85:** 133–146.

Wu-Scharf, D., B. Jeong, C. Zhang and H. Cerutti, 2000 Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. Science **290:** 1159–1162.

Xiong, Y., and T. H. Eickbush, 1990 Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:** 3353–3362.

Communicating editor: D. J. Grunwald