MOLECULAR AND CELLULAR BIOLOGY, June 1987, p. 2221–2230
0270-7306/87/062221-10$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 7, No. 6

The Site-Specific Ribosomal Insertion Element Type II of *Bombyx mori* (R2Bm) Contains the Coding Sequence for a Reverse Transcriptase-Like Enzyme

WILLIAM D. BURKE, CAROLYN C. CALALANG,† AND THOMAS H. EICKBUSH*
Department of Biology, River Campus, University of Rochester, Rochester, New York 14627

Received 29 December 1986/Accepted 12 March 1987

Two classes of DNA elements interrupt a fraction of the rRNA repeats of *Bombyx mori*. We have analyzed by genomic blotting and sequence analysis one class of these elements which we have named R2. These elements occupy approximately 9% of the rDNA units of *B. mori* and appear to be homologous to the type II rDNA insertions detected in *Drosophila melanogaster*. Approximately 25 copies of R2 exist within the *B. mori* genome, of which at least 20 are located at a precise location within otherwise typical rDNA units. Nucleotide sequence analysis has revealed that the 4.2-kilobase-pair R2 element has a single large open reading frame, occupying over 82% of the total length of the element. The central region of this 1,151-amino-acid open reading frame shows homology to the reverse transcriptase enzymes found in retroviruses and certain transposable elements. Amino acid homology of this region is highest to the mobile line 1 elements of mammals, followed by the mitochondrial type II introns of fungi, and the *pol* gene of retroviruses. Less homology exists with transposable elements of *D. melanogaster* and *Saccharomyces cerevisiae*. Two additional regions of sequence homology between L1 and R2 elements were also found outside the reverse transcriptase region. We suggest that the R2 elements are retrotransposons that are site specific in their insertion into the genome. Such mobility would enable these elements to occupy a small fraction of the rDNA units of *B. mori* despite their continual elimination from the rDNA locus by sequence turnover.

A fraction of the 28S ribosomal genes in several insect species is interrupted by segments of non-rDNA approximately 5 kilobases (kb) in length (see the review by Beckingham [3]). Based on the nucleotide sequences of their junction regions, these insertions have been divided into two classes. Type I elements have been reported in *Drosophila virilis* (34), *D. melanogaster* (7, 37), *Calliphora erythrocephala* (43), and *Bombyx mori* (12, 16). They interrupt the 28S rRNA gene at a location approximately two-thirds of the distance from the 5′ end of the gene and are flanked by a 14-base-pair (bp) duplication of rDNA sequences. Type II elements have only been reported in *D. melanogaster* (7, 37) and *B. mori* (12, 16). They interrupt the 28S gene 75 bp upstream of the type I elements and do not contain flanking duplications of rDNA sequences. In *D. melanogaster*, rDNA units containing either of these elements are transcribed at a significantly lower level than are the remaining rDNA units (20, 23, 28). To be consistent with the single-letter nomenclature frequently used to describe repetitive or transposable DNA elements and to avoid possible confusion with the type I and II intron sequences of fungal mitochondrial DNAs, we suggest that these elements be referred to as the R1 (type I) and R2 (type II) elements (R referring to rDNA units).

Disruption of the large rRNA genes has also been detected in noninsect species. In *Ascaris lumbricoides*, a 4.5-kb element interrupts a small fraction of the 26S rRNA genes at a position which corresponds to 34 bp upstream of the insect R1 element (2). Like the R1 element of insects, the *Ascaris* element is flanked by a 14-bp duplication of rDNA sequences. In *Tetrahymena* species, a 0.4-kb element interrupts all 26S rRNA genes at a position which corresponds to 3 bp 5′ of the insect R2 element (22, 50). Unlike the R1 and R2 elements of insects, the *Tetrahymena* rDNA disruption is a self-splicing intron, eliminated during the processing of rDNA transcripts (24). It is puzzling that there are so many cases of species containing insertions, all occurring within 78 bp in the large rRNA genes, particularly in *Ascaris* and insect species in which sequence turnover has neither fixed these insertions in all rDNA units nor eliminated them. This situation points to the existence of a mechanism of duplication and transposition to account for their maintenance.

In this report, we present the complete nucleotide sequence of an R2 element of *B. mori* (R2Bm). The element contains a 1,151-amino-acid open reading frame (ORF) which shows homology to the reverse transcriptases found in retroviruses and transposable elements. Greatest amino acid homology was found with the mammalian L1 elements (18, 27). Similar studies with the R1 element of *B. mori* are in progess.

MATERIALS AND METHODS

Identification and characterization of lambda clones containing R2Bm elements. Lambda clones containing rDNA units with R2 elements (B74, B98, and B131) or without this element (B108) were isolated from a Charon 4 library made from genomic DNA partially digested with *Eco*RI (11). The description of these clones and the sequence of their rDNA-R2 junction regions have been previously described (12). Clones B701, B702, and B703 were derived from a lambda library of partial *Sau*3A-I-digested genomic DNA cloned into Charon 35 (a gift from B. Hibner). Clones containing R2 elements were isolated from this library by hybridization to an 0.8-kb *Sst*I fragment isolated from the R2 element of B131. Hybridization data and restriction analysis indicated

* Corresponding author.
† Present address: University of Rochester Medical Center, Rochester, NY 14627.

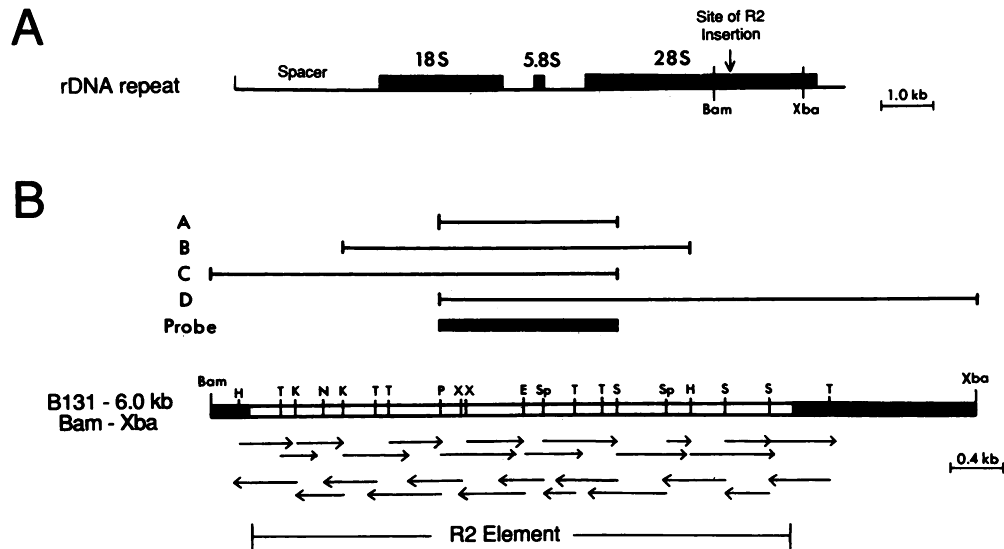

FIG. 1. Location and restriction map of the R2 elements within the rDNA units of *B. mori*. (A) Location of R2 within the 10.6-kb rDNA unit. Solid bars, rRNA genes; thin lines, transcribed and nontranscribed spacer regions. A 30-bp hidden break is found near the center of the 28S gene (15). The *Bam*HI and *Xba*I restriction sites used in subcloning the R2 elements are indicated. (B) Partial restriction map of the 6.0-kb *Bam*HI-*Xba*I fragment from lambda clone B131. Abbreviations for restriction sites used in the nucleotide sequence determinations are the following: H, *Hin*cII; T, *Taq*I; K, *Kpn*I; N, *Nar*I; P, *Pst*I; X, *Xho*I; E, *Eco*RI; Sp, *Sph*I; S, *Sst*I. Shown above the restriction map are the location and length of DNA fragments generated by R2 elements located in rDNA units. Lines A through D correspond to the restriction digestion in lanes A through D of Fig. 2. The solid bar correponds to the restriction fragment used as a probe in Fig. 2. Shown below the restriction map are the direction and extent of the sequence determinations. The complete sequence of both strands was determined.

that each of these clones contained a complete R2 element inserted into an otherwise typical rDNA unit.

Sequencing of R2Bm elements. A 6.0-kb *Bam*HI-*Xba*I fragment containing the complete R2 element and flanking ribosomal 28S sequences was subcloned from B131 into pUC13. After a detailed restriction map of this subclone was generated, the nucleotide sequence of the entire R2 element was determined by placing specific restriction fragments into mp18 and mp19 vectors (49). The nucleotide sequences of these fragments were determined by the dideoxy chain termination method (39). The specific restriction sites that were used and the direction of sequence determination from these sites are shown at the bottom of Fig. 1B. Both DNA strands of the element were completely sequenced, and all the restriction sites used in the subcloning were crossed on at least one strand. Nucleotide sequences from the lambda clones B701, B702, and B703 were obtained by subcloning fragments directly into the sequencing vectors. In the case of the 5′ junction region, a 0.76-kb *Bam*HI-*Kpn*I fragment was cloned into mp18 and mp19, and the junction sequence was determined on both strands. In the case of the 3′ junction region, a 1.6-kb *Sst*I-*Xba*I fragment was cloned into mp19, and the nucleotide sequence was determined from the internal *Sst*I site to 200 bp beyond the 3′ junction with the 28S gene. Finally, the internal segment of R2 corresponding to the beginning of the ORF was sequenced by subcloning the 0.36-kb *Kpn*I-*Kpn*I fragment into mp18 and determining its nucleotide sequence in both orientations.

RESULTS

Distribution of R2 elements within the genome. From our previously characterized collection of lambda clones containing uninterrupted or interrupted *B. mori* rDNA units (12, 25), clone B131 was selected for further analysis because it contained a complete R2 element. A 6.0-kb *Bam*HI-*Xba*I fragment of B131 containing the 4.2-kb R2 element, 0.3 kb of flanking 28S gene sequences at the 5′ junction of the element, and 1.5 kb of 28S gene sequences at the 3′ junction of the element was subcloned, and a detailed restriction map was generated (Fig. 1B).

Our first experiments were to determine the number and degree of sequence conservation of R2 elements within the *B. mori* genome by genomic blotting. An example of our results is shown in Fig. 2. The probe for the genomic blot in this figure was a 1.4-kb *Pst*I-*Sst*I fragment located at the center of the R2 element of B131. In lane A, the genomic DNA was digested with both *Pst*I and *Sst*I. Most of the DNA within the *B. mori* genome that hybridized with the R2 probe is located on a restriction fragment 1.4 kb in length. This fragment hybridized at an intensity equal to 20 to 25 copies per genome. In addition to this major hybridizing band, two fainter bands can also be seen in lane A. These bands hybridized at levels roughly equivalent to one copy per genome, suggesting that two of the R2 elements within the genome lost either the *Pst*I or *Sst*I cleavage sites. Longer exposures of the autoradiogram in Fig. 2 (not shown) revealed a series of even fainter bands in lane A which hybridized at levels significantly less than one copy per genome. These faintly hybridizing restriction fragments may represent either highly divergent copies of the R2 element or cross-hybridization of our R2 probes with another DNA element present in the genome of *B. mori*. (This second element is not R1, as we have never detected cross-hybridization between cloned copies of R1 and R2.)

The genomic DNA was also digested with *Kpn*I and *Hin*cII (Fig. 2, lane B). The dominant hybridizing fragment was 2.7 kb in length, identical to that found in the R2 element of B131 (Fig. 1B). In this digest, however, there were five additional fragments greater than 2.7 kb in length that hybridized at a level similar to one copy per genome and one 2.4-kb fragment that hybridized at a level equal to three to

four copies per genome. From the results of several combinations of restriction enzymes, most of these additional fragments correspond in the former case to the loss of the *Hin*cII site of the R2 element and in the latter case to the formation of a new *Hin*cII site in a subset of the R2 elements (data not presented). These data suggest that the R2 sequences correspond to a well-defined population of approximately 25 sequence elements within the *B. mori* genome. A low level of restriction polymorphism exists within the various copies of the element.

To determine what fraction of the R2 elements is present within rDNA units, genomic DNA was digested with several restriction enzymes that cleave once within the element and once within the rDNA unit immediately flanking the element. Figure 2, lane C, corresponds to a *Bam*HI-*Sst*I double digest of genomic DNA. The major hybridizing fragment generated by the genomic DNA was 3.1 kb in length, identical to that generated by the B131 subclone. This result indicates that most of the R2 elements are located approximately 0.3 kb downstream of the *Bam*HI site within the 28S gene. A similar result was obtained by using a conserved restriction site located within the rDNA unit 3′ of the R2 elements. Figure 2, lane D, corresponds to a *Pst*I-*Xba*I double digest of genomic DNA. As expected, the predominant band observed, 4.1 kb, was identical to that predicted for R2 elements inserted 0.3 kb downstream of the *Bam*HI site of the 28S gene. Of the five more faintly hybridizing bands seen in lane C and the six bands seen in lane D of the genomic blot, only two can be explained as resulting from variation in either the *Sst*I or the *Pst*I cleavage sites within the element (no variation exists in the *Bam*HI or *Xba*I sites of the 28S gene). Thus, there appear to be five copies of R2 elements not present within the rDNA units. Additional experiments with results similar to those shown in Fig. 2 have been conducted with probes from both the 5′ and 3′ ends of the R2 element and with enzymes that cleave at more distant locations within the rDNA units (data not shown). In both cases, these blots have indicated that approximately 25 copies of the R2 element are present in *B. mori* and that approximately 20 of these copies are inserted 0.3 kb downstream of the *Bam*HI site within the 28S rRNA gene of rDNA units. Since an estimated 240 rDNA units are present within the *B. mori* genome (17), approximately 9% contain R2 elements. This percentage is slightly higher than the 7% we have previously estimated from variant restriction fragments seen on genomic blots probed with 28S gene sequences (25).

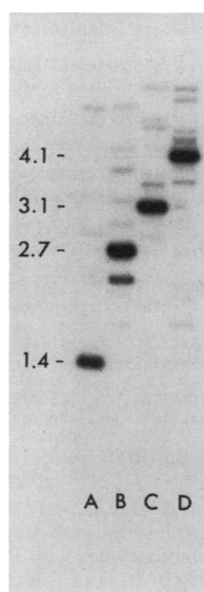

FIG. 2. Genomic blot of *B. mori* DNA probed with R2 sequences. For each lane, 3 μg of DNA was digested, fractionated on a 1.0% agarose gel, and transferred to nitrocellulose. The blot was hybridized with a nick-translated 1.4-kb *Pst*I-*Sst*I fragment isolated from the R2 element of clone B131 (Fig. 1B). Hybridization was conducted at 65°C in 0.6 M saline solution (0.6 M NaCl, 0.12 M Tris hydrochloride [pH 8], 4 mM EDTA), and final washes were at 65°C in 0.015 M saline (0.015 M NaCl, 0.003 M Tris hydrochloride [pH 8], 0.1 mM EDTA). The restriction endonucleases used were *Pst*I + *Sst*I (lane A), *Kpn*I + *Hin*cII (lane B), *Bam*HI + *Sst*I (lane C), and *Pst*I + *Xba*I (lane D). In each lane, the predominant hybridizing fragment is that predicted from the restriction map of the 6.0-kb *Bam*HI-*Xba*I subclone of clone B131 (Fig. 1B).

Sequence of R2-rDNA boundaries. In a previous report (12), we demonstrated that minor differences could be found in the sequence of the 5′ R2-rDNA junction region from three independent isolates of the R2 element (clones B74, B98, and B131). Because only one of these clones contained a complete R2 element, the extent of sequence variation at the 3′ junction could not be determined. To obtain a better understanding of this variation at R2 junctions with 28S genes, we isolated three additional lambda clones containing R2 elements. These new clones (B701, B702, and B703) were purified from a *Sau*3A-I genomic library of *B. mori* (a gift from B. Hibner), whereas the original clones were obtained from a partial *Eco*RI library (11). Each of these new clones contained an intact R2 element located within an otherwise typical rDNA repeat. The sequences of all junction regions from the six lambda clones are shown in Fig. 3 at the beginning and end of the complete sequence of the element from B131.

The variation detected in the 5′ R2-28S junctions of different copies is of three types, each type represented by two sequenced examples. In clones B131 and B702, there is a 24-bp duplication of the 28S gene sequence (Fig. 3). The two copies of this repeat are separated by 1 nucleotide. Clones B98 and B703 contain no duplication of the 28S gene sequence; instead, there is a 1-nucleotide deletion. This deleted base, an adenine, corresponds to the 1 nucleotide separating the DNA duplication in clones B131 and B702. Thus, the 5′ end of clones B98 and B703 could be generated from rDNA units similar to B131 and B702 by recombination between the 24-nucleotide repeats. Finally, in clones B74 and B701, there is neither a deletion nor a duplication of 28S sequences. Because of this variation at the 5′ boundary of the R2 element, the exact junction of the element with the 28S gene cannot be unambiguously defined. We believe the R2 element should be defined as starting with the sequence GGCGAT (Fig. 3). With the R2 element defined in this manner, all sequence variations that are detected at the 5′ junction of different copies are located within the 28S gene and not within the R2 element itself. Thus, for B131 and B702, the DNA duplication is from nucleotides −27 to −4, whereas in B98 and B701 the deletion is of nucleotide −3. As will be discussed in greater detail below, we believe R2 is a mobile DNA element. Therefore, some of the variants we have detected at the target site for the different copies of R2 may be intermediates in the insertion of the element into the 28S genes and thus are important clues to the mechanism and possible sequence specificity of this process.

In the case of the 3′ junction of the R2 element, four copies of the element have been sequenced (clones B74 and B98

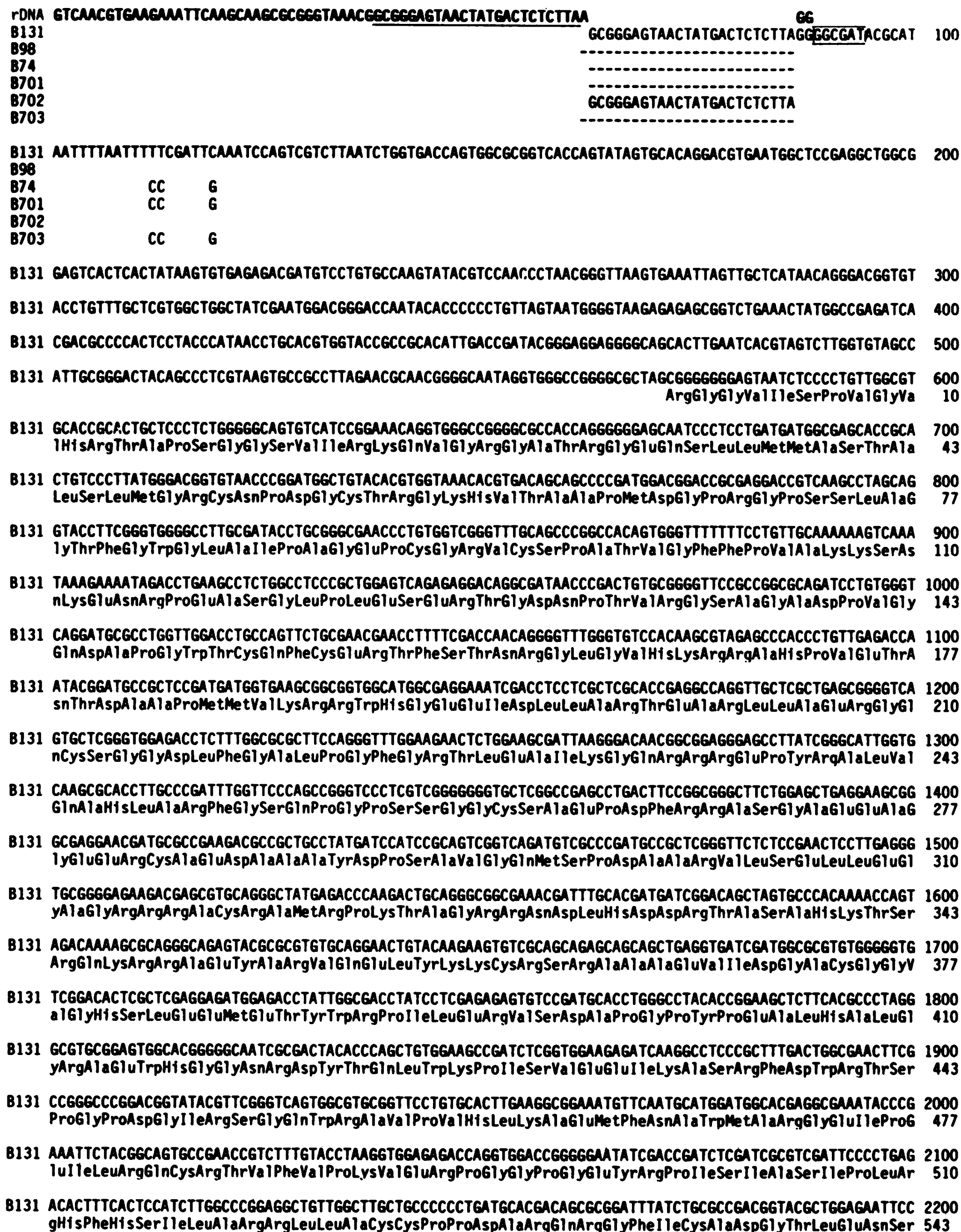

FIG. 3. Nucleotide sequence of R2. The complete nucleotide sequence from the R2 element of clone B131 is shown together with the predicted amino acid sequence of the ORF. The flanking 28S gene sequences determined from an uninserted rDNA unit (clone B108 [12]) are shown in boldface type. Also presented are sequences of the junction regions from additional lambda clones with R2 insertions. Individual nucleotides from these additional clones are only presented where they differ from the R2 element of B131 or from the rRNA sequences of an uninserted rDNA unit. Deleted nucleotides are indicated (−). All sequences are numbered from a *Hin*cII site located within the flanking 28S region, which was used in the sequence determination. Underlined nucleotides at the 5′ junction of the R2 element are nucleotides duplicated in clones B131 and B702; boxed nucleotides at the 5′ and 3′ junctions are repeated sequences near the boundaries of all R2 elements. Ter, Termination.

end at the *Eco*RI site within the R2 element [see Fig. 1 of reference 12]). All the copies end with the sequence GAAAA (Fig. 3). Thus, unlike the 5′ junction, no nucleotide variation was detected at the 3′ junction of different copies of R2.

Finally, a small degree of nucleotide variation was detected within the internal sequences of the R2 element. Clones B74, B701, and B703 contain three substitutions clustered approximately 25 bp from the 5′ end of the element. In the case of the 3′ end of the element, B701 contains one substitution 75 bp from the 3′ junction and a 3-bp deletion 160 bp from the 3′ junction. This low level of sequence variation within the different cloned copies of the R2 element is consistent with the restriction polymorphism seen in the genomic blot assays (Fig. 2).

```
B131 GCAGTACTGGACGCGGTGCTTGGGGATAGCAGGAAGAAGCTGCGGGAATGTCACGTGGCGGTGCTAGACTTCGCCAAGGCATTTGACACAGTGTCTCACG 2300
     AlaValLeuAspAlaValLeuGlyAspSerArgLysLysLeuArgGluCysHisValAlaValLeuAspPheAlaLysAlaPheAspTyrValSerHisG  577

B131 AGGCACTTGTCGAATTGCTGAGGTTGAGGGGCATGCCCGAACAGTTCTGCGGCTACATTGCTCACCTATACGATACGGCGTCCACCACCTTAGCCGTGAA 2400
     luAlaLeuValGluLeuLeuArgLeuArgGlyMetProGluGlnPheCysGlyTyrIleAlaHisLeuTyrAspThrAlaSerThrThrLeuAlaValAs  610

B131 CAATGAAATGAGCAGCCCTGTAAAAGTGGGACGAGGGGTTCGTCAAGGGGACCCTCTGTCGCCGATACTCTTCAACGTGGTGATGGACCTCATCCTGGCT 2500
     nAsnGluMetSerSerProValLysValGlyArgGlyValArgGlnGlyAspProLeuSerProIleLeuPheAsnValValMetAspLeuIleLeuAla  643

B131 TCCCTGCCGGAGAGGGTCGGGTATAGGTTGGAGATGGAACTCGTGTCCGCTCTGGCCTATGCTGACGACCTAGTCCTGCTTGCGGGGTCGAAGGTAGGGA 2600
     SerLeuProGluArgValGlyTyrArgLeuGluMetGluLeuValSerAlaLeuAlaTyrAlaAspAspLeuValLeuLeuAlaGlySerLysValGlyM  677

B131 TGCAGGAGTCCATCTCTGCTGTGGACTGTGTCGGTAGGCAGATGGGCCTACGCCTGAATTGCAGGAAAAGCGCGGTTCTGTCTATGATACCGGATGGCCA 2700
     etGlnGluSerIleSerAlaValAspCysValGlyArgGlnMetGlyLeuArgLeuAsnCysArgLysSerAlaValLeuSerMetIleProAspGlyHi  710

B131 CCGCAAGAAGCATCACTACCTGACTGAGCGAACCTTCAATATTGGAGGTAAGCCGCTCAGGCAGGTGAGTTGTGTTGAGCGGTGGCGATATCTTGGTGTC 2800
     sArgLysLysHisHisTyrLeuThrGluArgThrPheAsnIleGlyGlyLysProLeuArgGlnValSerCysValGluArgTrpArgTyrLeuGlyVal  743

B131 GATTTTGAGGCCTCTGGATGCGTGACATTAGAGCATAGTATCAGTAGTGCTCTGAATAACATCTCAAGGGCACCTCTCAAACCCCAACAGAGGTTGGAGA 2900
     AspPheGluAlaSerGlyCysValThrLeuGluHisSerIleSerSerAlaLeuAsnAsnIleSerArgAlaProLeuLysProGlnGlnArgLeuGluI  777

B131 TTTTGAGAGCTCATCTGATTCCGAGATTCCAGCACGGTTTTGTGCTTGGAAACATCTCGGATGACCGATTGAGAATGCTCGATGTCCAAATCCGGAAAGC 3000
     leLeuArgAlaHisLeuIleProArgPheGlnHisGlyPheValLeuGlyAsnIleSerAspAspArgLeuArgMetLeuAspValGlnIleArgLysAl  810

B131 AGTCGGACAGTGGCTAAGGCTACCGGCGGATGTGCCCAAGGCATATTATCACGCCGCAGTTCAGGACGGCGGCTTAGCGATCCCATCGGTGCGAGCGACC 3100
     aValGlyGlnTrpLeuArgLeuProAlaAspValProLysAlaTyrTyrHisAlaAlaValGlnAspGlyGlyLeuAlaIleProSerValArgAlaThr  843

B131 ATCCCGGACCTCATTGTGAGGCGTTTCGGGGGGCTCGACTCGTCACCATGGTCAGTGGCAAGAGCCGCCGCCAAATCTGATAAGATTCGTAAGAAACTGC 3200
     IleProAspLeuIleValArgArgPheGlyGlyLeuAspSerSerProTrpSerValAlaArgAlaAlaAlaLysSerAspLysIleArgLysLysLeuA  877

B131 GGTGGGCCTGGAAACAGCTCCGCAGGTTCAGCCGTGTTGACTCCACAACGCAACGACCATCTGTGCGCTTGTTTTGGCGAGAACATCTGCATGCATCTGT 3300
     rgTrpAlaTrpLysGlnLeuArgArgPheSerArgValAspSerThrThrGlnArgProSerValArgLeuPheTrpArgGluHisLeuHisAlaSerVa  910

B131 TGATGGACGCGAACTTCGCGAATCCACACGCACCCCGACATCCACAAAGTGGATTAGGGAGCGATGCGCGCAGATAACCGGACGGGACTTCGTGCAGTTC 3400
     lAspGlyArgGluLeuArgGluSerThrArgThrProThrSerThrLysTrpIleArgGluArgCysAlaGlnIleThrGlyArgAspPheValGlnPhe  943

B131 GTGCACACTCATATCAACGCCCTCCCATCCCGCATTCGCGGATCGAGAGGGCGTAGAGGTGGGGGTGAGTCTTCGTTGACCTGCCGTGCTGGTTGCAAGG 3500
     ValHisThrHisIleAsnAlaLeuProSerArgIleArgGlySerArgGlyArgArgGlyGlyGlyGluSerSerLeuThrCysArgAlaGlyCysLysV  977

B131 TTAGGGAGACGACGGCTCACATCCTACAACAGTGTCACAGAACACACGGCGGCCGGATTCTACGACACAACAAGATTGTATCTTTCGTGGCGAAAGCCAT 3600
     alArgGluThrThrAlaHisIleLeuGlnGlnCysHisArgThrHisGlyGlyArgIleLeuArgHisAsnLysIleValSerPheValAlaLysAlaMe 1010

B131 GGAAGAGAACAAGTGGACGGTTGAGCTGGAGCCGAGGCTACGAACATCGGTTGGTCTCCGTAAGCCGGATATTATCGCCTCCAGGGATGGTGTCGGAGTG 3700
     tGluGluAsnLysTrpThrValGluLeuGluProArgLeuArgThrSerValGlyLeuArgLysProAspIleIleAlaSerArgAspGlyValGlyVal 1043

B131 ATCGTGGACGTGCAGGTGGTCTCGGGCCAGCGATCGCTTGACGAGCTCCACCGTGAGAAACGTAATAAATACGGGAATCACGGGGAGCTGGTTGAGTTGG 3800
     IleValAspValGlnValValSerGlyGlnArgSerLeuAspGluLeuHisArgGluLysArgAsnLysTyrGlyAsnHisGlyGluLeuValGluLeuV 1077

B131 TCGCAGGTAGACTAGGACTTCCGAAAGCTGAGTGCGTGCGAGCCACTTCGTGCACGATATCTTGGAGGGGAGTATGGAGCCTGACTTCTTATAAGGAGTT 3900
     alAlaGlyArgLeuGlyLeuProLysAlaGluCysValArgAlaThrSerCysThrIleSerTrpArgGlyValTrpSerLeuThrSerTyrLysGluLe 1110

B131 AAGGTCCATAATCGGGCTTCGGGAACCGACACTACAAATCGTTCCGATACTGGCGTTGAGAGGTTCACACATGAACTGGACCAGGTTCAATCAGATGACG 4000
     uArgSerIleIleGlyLeuArgGluProThrLeuGlnIleValProIleLeuAlaLeuArgGlySerHisMetAsnTrpThrArgPheAsnGlnMetThr 1143

B131 TCCGTCATGGGGGGCGGCGTTGGTTGAGCCTTGCACAGTAGTCCAGCGGTAAGGGTGTAGATCAGGCCCGTCTGTTTCTCCCCCGGAGCTCGCTCCCTTG 4100
     SerValMetGlyGlyGlyValGlyTer

B131 GCTTCCCTTATATATTTTAACATCAGAAACAGACATTAAACATCTACTGATCCAATTTCGCCGGCGTACGGCCACGATCGGGAGGGTGGGAATCTCGGGG 4200
B701             ---
B702
B703

rDNA                                                                                TAGCCAAATGCCTCGTCATCTAATT 4300
B131 GTCTTCCGATCCTAATCCATGATGATTACGACCTGAGTCACTAAAGACGATGGCATGATGATCC[GGCGAT]GAAAA
B701 A
B702
B703

rDNA AGTGACGCGCATGAATGGATTAACGAGATTCCCACTGTCCCTATCTACTATCTAGCGAAACCACAGCCAAGGGAACGGGCTTGGGAGAATCAGCGGGGAA 4400
B701
B702
B703
```

FIG. 3—*Continued*

Sequence of a complete R2 element. The nucleotide sequence of the entire R2 element subcloned from B131 was determined on both strands as shown in Fig. 1B. The element is quite GC rich (57%), which is due almost entirely to an abundance of guanine (32%) on the strand shown in Fig. 3. In localized regions of the element, this preponderance of guanine on one strand exceeds 50%, giving rise to multiple runs of consecutive guanines 4 to 7 nucleotides in length. It is interesting that the region with the highest G+C content (67%) involves a duplication of the segment encoding the beginning of the ORF (see below). The 89-bp sequence from 549 to 637 is duplicated from 638 to 728. This duplication is not of recent origin since 15-nucleotide substitutions and two short insertion-deletion events have accumulated in the two copies. Each repeat contains an extremely guanine-rich region (19 of 30 bp) and cytosine-rich region (15 of 30 bp) which has significant potential for fold-back base pairing.

Perhaps the most striking feature derived from the complete sequence of an R2 element is the presence of an ORF occupying over 82% of the entire sequence. This 1,151-amino-acid ORF is in the same orientation as is transcription of the rRNA genes, starting 483 bp from the 5′ junction with the 28S gene and ending 248 bp from the 3′ junction. The

FIG. 4. Homology of amino acid sequences between the R2 ORF and putative RNA-dependent DNA polymerase ORFs of various origins. Conserved amino acid positions identified by Toh et al. (45, 46) are grouped into seven regions (a to g). The amino acid sequences are shown for the following: L1Md, line 1 of mice (27); L1Hs, line 1 of humans (18); L1Nc, line 1 of the slow loris (18); *oxi3-1*, first intron of the yeast mitochondrial cytochrome oxidase subunit 1 gene (4); *oxi3-2*, second intron of the yeast cytochrome oxidase subunit 1 gene (4); RSV, Rous sarcoma virus *pol* gene (40); MuLV, Moloney murine leukemia virus *pol* gene (42); HTLV, human T-cell leukemia type III virus *pol* gene (35). Amino acids that are identical between R2 and the other sequences are indicated by boldface type; chemically similar amino acids at the same position are underlined. Chemically similar amino acids are defined as pairs of residues belonging to the same group as follows (41): A, G, P, S, and T; D, E, N, and Q; H, R, and K; I, L, M, and V; F, W, and Y. Location of amino acids reported by Toh et al. (45, 46) as invariant (●) and chemically similar (○) in retroviruses and amino acids which were not identified by Toh et al. but which are identical or similar in all nine sequences compared in the figure (★) are indicated. Numbers in parentheses indicate the number of amino acids omitted from the figure.

protein that would be encoded by this ORF is highly charged (28% charged residues), with a ratio of basic to acidic residues of 1.75. The most abundant amino acid is Arg, corresponding to 11.2% of the total. The first methionine residue of this ORF is at position 38. While it is unusual to have the initiation of translation at a position this far into an ORF, no ATG start codons are present in alternate frames of the sequence before this in-frame start at position 38. To eliminate the possibility that B131 contained a sequence variation disrupting the normal start site for the ORF, we have sequenced this region of the element from three additional cloned copies of the R2 element. Clones B701, B702, and B703 are all identical in sequence to B131.

The ORF of R2 shows homology to reverse transcriptases. A number of transposable elements found in *D. melanogaster*, yeasts, and mammals contain one or two ORFs occupying nearly the entire length of the element. The elements are frequently referred to as retrotransposons because their largest ORFs exhibit homology to retroviral RNA-dependent DNA polymerase. The central region of the R2 ORF also shows significant sequence homology to the RNA-dependent DNA polymerases of retroviruses (35, 40, 42) (Fig. 4). Also included in this figure are sequences from several elements with the highest level of homology to R2Bm: three mammalian line 1 sequences (18, 27) and two yeast mitochondrial intron sequences which have been shown to exhibit homology to reverse transcriptase (4, 31). The locations of these regions of homology within the ORFs of each sequence are shown (see Fig. 6). While ORFs from the transposable elements 17.6, copia, 412, and gypsy of *D. melanogaster* (30, 32, 38, 51) and the Ty element of *Saccharomyces cerevisiae* (6) also exhibit homology to reverse transcriptase, they have not been included in these figures, because their homology to R2Bm is less than that of the retroviral sequences.

A comparison of retroviral *pol* genes and the large ORF of the transposable element 17.6 has allowed Toh et al. (45, 46) to identify a 175-amino-acid region common to all of these sequences in which 29 positions are either invariant or contain chemically similar amino acids. These conserved amino acids have been used to identify homology to reverse transcriptase in ORFs from a variety of sources (27, 31, 32). While exceptions to these invariant and similar positions have been found as the number of retroviral sequences increases (see, for instance, the human T-cell lymphotropic virus type III sequence in Fig. 4), these 29 positions remain useful diagnostic indicators for detecting sequence homology to reverse transcriptase-like enzymes. In Fig. 4, the conserved amino acid positions identified by Toh et al. (45, 46) have been grouped into seven regions labeled a to g. Those amino acid positions which are identical in most retroviruses and those positions with similar amino acids in most retroviruses are indicated. The R2Bm ORF contains the same amino acid in 11 of the 13 positions that were identified as identical in retroviruses. The nonconserved positions include an Arg substituted for a Pro in region d and a Thr substituted for a Gly in region g. R2Bm contains chemically similar amino acids in 10 of the 16 positions that were identified as similar among retroviral sequences.

We have also indicated in Fig. 4 all those amino acids in R2Bm that are conserved in one or more of the other sequences. In total, 69 of the possible 112 amino acid positions in regions a to g of R2Bm are conserved in at least one-half of the line 1, mitochondrial, and retroviral sequences. Indeed, we have indicated in Fig. 4 nine positions

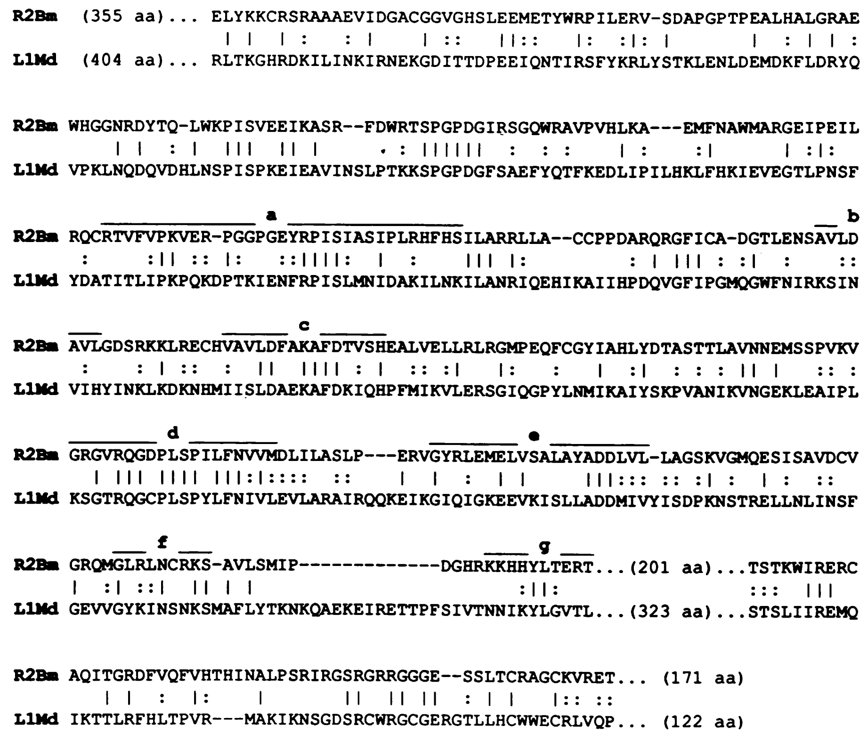

FIG. 5. Sequence comparison of the ORF of R2 and the large ORF of mouse L1 (L1Md). The sequence of mouse L1 was obtained from Loeb et al. (27). Identical amino acids are indicated by a vertical line, and chemically similar amino acids (defined in the legend to Fig. 4) are indicated by a colon. Segments labeled a to g correspond to the regions of homology to reverse transcriptases as presented in Fig. 4. Sequence homology for all regions presented is 26% (total number of identical amino acids divided by the total number of amino acid positions). aa, Amino acids.

in addition to those previously identified by Toh et al. that contain identical or similar amino acids in all nine sequences compared in the figure.

The R2 ORF shows highest homology to the ORF from L1 elements. Close inspection of the sequences in Fig. 4 reveals that the reverse transcriptase region of R2 has greater homology to the mammalian line 1 sequences than to either the mitochondrial or retroviral sequences. This greater homology can be scored in two ways. First, in the seven regions shown, L1 sequences share an average of 40 identical amino acid positions with R2 (range, 39 to 41 amino acids). This compares with 34 identical amino acids in *oxi3-1* and *oxi3-2* and an average of 29 identical amino acids for the retroviruses (range, 24 to 33 amino acids). Second, the distances separating the seven regions shown in Fig. 4 are highly similar in the R2 and L1 sequences, while in most cases they differ from those of the *oxi3* and retroviral sequences. In light of this greater homology of R2 and L1 sequences, we have conducted a detailed sequence comparison of L1 and R2 over the complete length of their ORFs. (L1 actually contains two partially overlapping ORFs; one is 461 amino acids in length and the other is 1,300 amino acids in length. Homology between R2 and L1 has only been detected with the second ORF.)

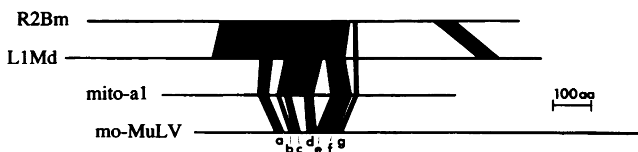

FIG. 6. Summary diagram indicating the regions of homology in the ORFs with putative reverse transcriptase enzymes. The horizontal lines correspond to the ORFs (amino-terminal ends at left), with homologous regions indicated by shading. The regions labeled a to g correspond to the reverse transcriptase regions presented in Fig. 4. Although they exhibit limited sequence homology, the segments between regions e and f are shaded because the length of this region is rigidly conserved in all four ORFs. Line 1 (L1Md) and R2 (R2Bm) share additional homology in the region preceding a to g and in a region near the carboxyl-terminal end of each ORF. mito-a1, First intron of the yeast mitochondrial *oxi-3* gene; mo-MuLV, Moloney murine leukemia virus *pol* gene; aa, amino acids.

The three highest regions of sequence homology between R2 and the large ORF of the mouse L1 sequence are shown in Fig. 5. The largest of these regions is the segment with homology to reverse transcriptase. This region is 240 amino acids in length in R2; 62 positions are identical in L1, corresponding to 26% homology. Note that considerable sequence homology is present throughout the region, not just in the seven regions conserved in all reverse transcriptase-like enzymes (labeled a to g).

The second region of homology between R2 and L1 is in the segment immediately preceding regions a to g. Thirty-two identical amino acids are found in 126 positions, corresponding to 25% homology. This region is not as conserved among retroviruses, corresponding to the 5′ end of the reverse transcriptase region and extending into the protease region in certain retroviruses (9, 26, 46).

The third segment of homology between R2 and L1 is found approximately midway between the reverse transcriptase region and the carboxyl-terminal end of the ORF (Fig. 6). In an area of 60 amino acids, 16 positions were identical between R2 and L1, corresponding to 27% homology. This

region corresponds to CS1, one of the two regions in the L1 element that is most conserved among mammals (5). In retroviruses, this region of the *pol* ORF is believed to encode an endonuclease (9, 48). We have not detected any homology between this conserved region of R2 and L1 and the putative endonuclease region of retroviruses (21, 46). While speculation as to the function of these two additional regions of homology between R2 and L1 cannot be made at this time, we believe these regions are important in the propagation of the elements because the extent of their conservation is similar to that of the reverse transcriptase region.

DISCUSSION

Our finding of a long ORF occupying over 82% of the coding capacity of the R2 element suggests that this element is under selective pressure for protein production. The amino acid sequence homology of this ORF to reverse transcriptase further suggests that R2 may be a mobile element that migrates via an RNA intermediate. Although the sequence of a complete type II element from *D. melanogaster* has not been determined, several similarities indicate that these elements are homologous to the *B. mori* element (12, 16): (i) their identical location within the 28S gene, (ii) the absence of target site duplication, and (iii) a short region of DNA sequence homology at their 3' junction. The mechanism of R2 insertion into the 28S genes exhibits two differences between these species. First, the 3' end of the *Drosophila* elements contain a poly(A) tail from 13 to 22 nucleotides in length (7, 37). The poly(A) tail at the 3' end of the R2Bm element is always 4 nucleotides in length. This precise 3' end of the R2Bm element may be the result of a specific cleavage of a longer poly(A) tail. Second, many of the *Drosophila* R2 elements are not full-length copies. These shortened copies contain deletions starting at their 5' end and extending to several preferred positions within the element (7, 37). We have not found such deletions within the R2 elements of *B. mori*. It will be of interest to determine if the *Drosophila* R2 elements contain an ORF and if so whether they contain regions of conserved sequence with the R2Bm ORF not shared with other transposable elements.

The high degree of sequence homogeneity of ribosomal repeats is believed to be a result of a high rate of unequal crossover or gene conversion, which leads to the elimination on fixation of new variants. These turnover mechanisms can also account for the spread of an insertion element as an rDNA variant. If such an element is deleterious to the expression of the rRNA genes, negative selection pressure would favor their elimination. In unusual circumstances in which reduced expression of ribosomal genes can be interpreted as conferring a selective advantage (8, 44), the maintenance of insertions might be favored. It seems unlikely that such a selection mechanism could have maintained these elements over the long interval since the divergence of *Drosophila* and *Bombyx*. The classification of R2 as a mobile element provides an additional property which may be the key factor in explaining its maintenance. In this model, the observed number of ribosomal genes with R2 elements is the result of an ongoing process involving the insertion of new copies into rDNA units and their expansion and elimination influenced by unequal crossovers, gene conversions, and selection. Similar suggestions that the transposability of the type I and type II elements aid their maintenance within the ribosomal locus have been previously offered (10, 36).

If R2Bm is a mobile element, it has at least two characteristics that are not typically associated with retroviruses and retrotransposons. First, R2 elements do not contain the long terminal repeats (LTRs) found in all retroviruses and most transposable elements. It is interesting that the elements which exhibit the highest homology to R2, the L1 elements of mammals, also lack LTRs (18, 27). This suggests a closer relationship between R2 and L1 than between R2 and other retrotransposons. The recent publication of the sequence of the I factor in *D. melanogaster* and its homology to L1 (13) greatly strengthens such a grouping of non-LTR elements. A detailed comparison of these three elements with a fourth non-LTR element, the R1 sequence of *B. mori*, will be presented in a subsequent report (Y. Xiong and T. H. Eickbush, manuscript in preparation).

The LTRs of retroviruses and transposable elements contain the promoter for RNA transcription. This enables the replication of these elements via an RNA intermediate with no net loss of sequence information (see the review by Varmus [47]). In the case of the L1 elements, a model has been proposed in which RNA transcription originates in a series of tandem duplications at the 5' end of the element (27). This model permits propagation of the element with no net loss of sequence information. The only repeat within R2Bm is not an exact duplication and is located over 500 bp into the element; thus, it is unlikely that this repeat can serve as a promoter for RNA transcription. Rather, we suggest that the R2 element does not contain its own promoter. Because R2 is situated in the same orientation as that of the rRNA genes, a full-length RNA molecule can be obtained by processing a PolI transcript of the rDNA unit. As a consequence, the R2 element can satisfy a fundamental requirement of a retrotransposon, production of a full-length transcript without the duplication of a region containing the promoter. While nothing is known of how the R2 element might be processed from the rRNA transcript, it may be significant that the sequence GGCGAT, defined as the 5' boundary of R2, is repeated immediately upstream of the short poly(A) tail at the 3' boundary (Fig. 3). This sequence may play an important role in the processing of the element. Consistent with this suggestion, this sequence is the major portion of the only region of sequence homology found at the 3' ends of R2 elements from *B. mori* and *D. melanogaster*.

A second feature of the R2 element that is not typically associated with retroviruses or transposable elements is the high degree of site specificity of its insertion into the genome. Of the estimated 25 copies of the R2 element present in the *B. mori* genome, approximately 20 are inserted at an identical location, occupying approximately 9% of the 28S genes. In the case of *D. melanogaster*, no copies of R2 elements are found outside the rDNA units. We are isolating the extraribosomal copies of the R2 element from *B. mori* to determine if they correspond to elements inserted into non-28S sequences or if they are associated with fragmented copies of the rDNA unit discarded from the rDNA loci.

Although preferred insertion sites have been reported with certain higher eucaryotic transposable elements (14, 33), these preferred sites are so short or imprecisely defined that they are present in large numbers throughout the genome. Thus, the mobile elements are distributed essentially randomly throughout the genome. The only example of a DNA element that is known to transpose with a specificity as high as that proposed for R2 is the transposition of the intron located in the large rRNA gene of certain yeast mitochondria (19, 29). The strict sequence specificity of this intron is believed to be a result of an endonuclease encoded by the 240-amino-acid ORF of the intron. We have not detected sequence homology between this ORF and R2. However,

sequence homology does exist in the reverse transcriptase regions of the ORFs of R2 and the type II introns of yeast mitochondria (Fig. 4 and 6). This homology is not as high as that shared by L1 elements and R2, but it is significantly higher than the homology of R2 with retroviruses and copialike transposable elements. Although the transposition of the type II introns of yeast mitochondria has not been reported, it has recently been demonstrated that the integration of a mitochondrial element related to the type II introns, the Mauriceville and Varkud mitochondrial plasmid DNAs of *Neurospora* spp., can occur via an RNA intermediate (1).

Finally, our classification of R2 as a site-specific mobile element raises the question of whether the evolution of site specificity has been a rare occurrence or whether there might be a class of similar elements. Such a class would be unlikely to be detected in the manner that most transposable elements have been discovered, i.e., as the cause of newly arising mutations. In any case, the ribosomal locus has several properties that appear to make it a uniquely suitable niche for the evolution of a site-specific mobile element. First, since the ribosomal genes are actively transcribed in essentially all tissues at all times, the accessibility of the target site for both insertion and transcription is highly favorable. Indeed, as we have suggested above, the R2 element probably relies upon the rRNA promoter for its transcription. Second, because rRNA genes are highly repeated, a site-specific mobile element retains the advantage of multiple target sites for insertion, and apparently a significant fraction of these sites can be occupied at a tolerable cost to the host. Third, since the DNA sequence of ribosomal genes is evolutionarily conserved, any mobile element which has developed site specificity for these genes is not precluded from interspecies mobility. In fact, it appears that two distinct elements currently occupy this niche in the *B. mori* genome. Preliminary studies of the R1 element of *B. mori* indicate that it also has properties of a retrotransposon (Xiong and Eickbush, in preparation).

ACKNOWLEDGMENTS

We thank Chung-I Wu for discussions and Stuart Horowitz for assistance in computer homology searches. We thank Yue Xiong for discussions, assistance in the genomic blotting experiments, and comments on the manuscript.

The investigation was supported in part by a Public Health Service grant from the National Institutes of Health and a deKiewiet Fellowship to C.C.C. from the University of Rochester.

LITERATURE CITED

1. **Akins, R. A., R. L. Kelley, and A. M. Lambowitz.** 1986. Mitochondrial plasmids of *Neurospora*: integration into mitochondrial DNA and evidence for reverse transcription in mitochondria. Cell **47**:505–516.
2. **Back, E., E. Van Meir, F. Muller, D. Schaller, H. Neuhaus, P. Aeby, and H. Tobler.** 1984. Intervening sequences in the ribosomal RNA genes of *Ascaris lumbricoides*: DNA sequences at the junctions and genomic organization. EMBO J. **3**:2523–2529.
3. **Beckingham, K.** 1982. Insect rDNA, p. 205–269. *In* H. Busch and L. Rothblum (ed.), The cell nucleus, vol. X, part A. Academic Press, Inc., New York.
4. **Bonitz, S. G., G. Coruzzi, B. E. Thalenfeld, A. Tzagoloff, and G. Macino.** 1980. Assembly of the mitochondrial membrane system: structure and nucleotide sequence of the gene coding for subunit 1 of yeast cytochrome oxidase. J. Biol. Chem. **255**:11927–11941.
5. **Burton, F. H., D. D. Loeb, C. F. Voliva, S. L. Martin, M. H. Edgell, and C. A. Hutchison III.** 1986. Conservation throughout Mammalia and extensive protein-encoding capacity of the highly repeated DNA long interspersed sequence one. J. Mol. Biol. **187**:291–304.
6. **Clara, J., and P. Farabaugh.** 1985. Nucleotide sequence of a yeast Ty element: evidence for a novel mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82**:2829–2833.
7. **Dawid, I. B., and M. L. Rebbert.** 1981. Nucleotide sequence at the boundaries between gene and insertion regions in the rDNA of *Drosophila melanogaster*. Nucleic Acids Res. **9**:5011–5020.
8. **DeSalle, R., J. Slightom, and E. Zimmer.** 1986. The molecular through ecological genetics of abnormal abdomen. II. Ribosomal DNA polymorphism is associated with the abnormal abdomen syndrome in *Drosophila mercatorium*. Genetics **112**:861–875.
9. **Dickson, C., R. Eisenman, R. H. Fan, E. Hunter, and N. Teich.** 1982. Protein biosynthesis and assembly, p. 513–648. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Dover, G., and E. Coen.** 1981. Springcleaning ribosomal DNA: a model for multigene evolution. Nature (London) **290**:731–732.
11. **Eickbush, T. H., and F. C. Kafatos.** 1982. A walk in the chorion locus of *Bombyx mori*. Cell **29**:633–643.
12. **Eickbush, T. H., and B. Robins.** 1985. *Bombyx mori* 28S genes contain insertion elements similar to the type I and II elements of *Drosophila melanogaster*. EMBO J. **4**:2281–2285.
13. **Fawcett, D. H., C. K. Lister, E. Kellet, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. Cell **47**:1007–1015.
14. **Freund, R., and M. Meselson.** 1984. Long terminal repeat nucleotide sequence and specific insertion of the gypsy transposon. Proc. Natl. Acad. Sci. USA **81**:4462–4464.
15. **Fujiwara, H., and H. Ishikawa.** 1986. Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects: implication based on structural studies. Nucleic Acids Res. **14**:6393–6401.
16. **Fujiwara, H., T. Ogura, N. Takada, N. Miyajima, H. Ishikawa, and H. Maekawa.** 1984. Introns and their flanking sequences of *Bombyx mori* rDNA. Nucleic Acids Res. **12**:6861–6869.
17. **Gage, L. P.** 1974. Polyploidization of the silkgland of *Bombyx mori*. J. Mol. Biol. **86**:97–108.
18. **Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki.** 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature (London) **321**:625–628.
19. **Jacquier, A., and B. Dujon.** 1985. An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. Cell **41**:383–394.
20. **Jamrich, M., and O. L. Miller.** 1984. The rare transcripts of interrupted rRNA genes in *Drosophila melangoster* are processed or degraded during synthesis. EMBO J. **3**:1541–1545.
21. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83**:7648–7652.
22. **Kan, N. C., and J. G. Gall.** 1982. The intervening sequence of the ribosomal RNA gene is highly conserved between two *Tetrahymena* species. Nucleic Acids Res. **10**:2809–2822.
23. **Kidd, S. J., and D. M. Glover.** 1981. *Drosophila melanogaster* ribosomal DNA containing type II insertions is variably transcribed in different strains and tissues. J. Mol. Biol. **151**:645–662.
24. **Kruger, K., P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech.** Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. Cell **31**:147–157.
25. **Lecanidou, R., T. H. Eickbush, and F. C. Kafatos.** 1984. Ribosomal DNA genes of *Bombyx mori*: a minor fraction of the repeating units contain insertions. Nucleic Acids Res. **12**:4703–4713.
26. **Levin, J. G., S. C. Hu, A. Rein, L. I. Messer, and B. I. Gerwin.** 1984. Murine leukemia virus mutant with a frameshift in the reverse transcriptase coding region: implications for *pol* gene

structure. J. Virol. **51**:470–478.

27. **Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchison III.** 1986. The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol. Cell. Biol. **6**:168–182.
28. **Long, E. O., and I. B. Dawid.** 1979. Expression of ribosomal DNA insertions in *Drosophila melanogaster*. Cell **18**:1185–1196.
29. **Macreadie, I. G., R. M. Scott, A. R. Zinn, and R. A. Butow.** 1985. Transposition of an intron in yeast mitochondria requires a protein encoded by that intron. Cell **41**:395–402.
30. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6**:1129–1134.
31. **Michel, F., and B. F. Lang.** 1985. Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. Nature (London) **316**:641–643.
32. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5**:1630–1638.
33. **O'Hara, K., and G. M. Rubin.** 1983. Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**:25–35.
34. **Rae, P. M. M., B. D. Kohorn, and R. P. Wade.** 1980. The 10 kb *Drosophila virilis* intervening sequence is flanked by a direct repeat of 14 base pairs of coding sequence. Nucleic Acids Res. **8**:3491–3505.
35. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papis, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
36. **Roiha, H., and D. M. Glover.** 1981. Duplicated rDNA sequences of variable lengths flanking the short type I insertions in the rDNA of *Drosophila melanogaster*. Nucleic Acids Res. **9**:5521–5532.
37. **Roiha, H., J. R. Miller, L. C. Woods, and D. M. Glover.** 1981. Arrangements and rearrangements of sequences flanking the two types of rDNA insertion in *D. melanogaster*. Nature (London) **290**:749–753.
38. **Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki.** 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature (London) **312**:659–661.
39. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci USA **74**:5463–5467.
40. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of the Rous sarcoma virus. Cell **32**:853–869.
41. **Schwartz, R. M., and M. O. Dayhoff.** 1978. Matrices for detecting distant relationships, p. 353–358. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and stucture, vol. 5. National Biomedical Research Foundation, Washington, D.C.
42. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
43. **Smith, V. L., and K. Beckingham.** 1984. The intron boundaries and flanking rRNA coding sequences of *Calliphora erythrocephala* rDNA. Nucleic Acids Res. **12**:1707–1724.
44. **Templeton, A. R., T. J. Crease, and F. Shah.** 1985. The molecular through ecological genetics of abnormal abdomen in *Drosophila mercatorium*. I. Basic genetics. Genetics **111**:805–818.
45. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral transcriptase and putative polymerases of hepatitus B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
46. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and pol gene product of Moloney murine leukemia virus. EMBO J. **4**:1267–1272.
47. **Varmus, H. E.** 1983. Retroviruses, p. 411–503. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., Orlando, Fla.
48. **Varmus, H. E., and R. Swanstrom.** 1982. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
50. **Wild, M. A., and M. Sommer.** 1980. Sequence of the ribosomal RNA gene intron from *Tetrahymena*. Nature (London) **283**:693–694.
51. **Yuki, S., S. Ishimura, S. Inouye, and K. Saigo.** 1986. Identification of genes for reverse transcriptase-like enzymes in two *Drosophila* retrotransposons, 412 and gypsy: a rapid detection method of reverse transcriptase genes using YXDD box probes. Nucleic Acids Res. **14**:3017–3030.