Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 8401–8404, October 1991
Genetics

Retroviral-like element in a marine invertebrate

(retrotransposon/retrovirus/long terminal repeat/echinoderm/mobile elements)

MARK S. SPRINGER*, ERIC H. DAVIDSON, AND ROY J. BRITTEN

Division of Biology, California Institute of Technology, Pasadena, CA 91125

Contributed by Roy J. Britten, June 17, 1991

ABSTRACT **Retroviral-like elements (RL elements) include retroviruses and long terminal repeat (LTR)-containing retrotransposons. We report the presence of sea urchin RL elements (termed *SURL*) in eight species of sea urchins and find that these RL elements belong to several subfamilies. The complete DNA sequence of one *SURL* element in *Tripneustes gratilla* is 5266 base pairs long, including 254-nucleotide-long identical long terminal repeats (LTRs). It contains a single open reading frame nearly 4 kilobases long including the *gag* and *pol* genes. Comparison of conserved DNA sequences of RL elements from different sea urchin species indicates that active elements have been inserting copies into echinoid genomes for at least 200 million years.**

Mobile genetic elements have strongly affected the evolution of the eukaryotic genome, and retrotransposons have been of central importance among them. Retrotransposons that contain long terminal repeats (LTRs) have a striking resemblance to the proviral form of vertebrate retroviruses and include coding regions for most of the same genes in the same order (1). Both encode a reverse transcriptase (RT), and DNA copies of transcripts are inserted into the host genome. Together, LTR-containing retrotransposons and retroviruses are conveniently termed "retroviral-like elements" (RL elements) (2).

RL elements are known in a wide variety of organisms, including yeast (3–5), dipterans (6), lepidopterans (7), gymnosperms (8), angiosperms (9–11), and vertebrates, where they are recognized as retroviruses or endogenous retroviruses and sometimes as retrotransposons (12). RL elements are apparently widely distributed among echinoids. We have observed the RT gene coding region of RL elements in the DNA of all sea urchin species that we have examined, including one cidaroid (*Eucidaris tribuloides*) and seven euechinoids (*Strongylocentrotus purpuratus, Strongylocentrotus franciscanus, Strongylocentrotus drobachiensis, Tripneustes gratilla, Lytechinus pictus, Lytechinus variegatus*, and *Arbacia punctulata*).

In early work on long, interspersed, repeated DNA of sea urchins, electron micrographic examination of reassociated DNA heteroduplexes revealed elements (called cs2108) that were about 5.2 kilobases (kb) long and had an interesting family and subfamily pattern (13). The elements are now named sea urchin RL elements (*SURL* elements). There are ≈400 copies of the RT gene coding region interspersed in the genome of *S. purpuratus* (13, 14). Cloned elements are named as in the example SURL1-21 (Sp), where the number 1 defines the *SURL* subfamily, 21 is an arbitrary label for the individual clone, and the letters in parentheses are the acronym of the species from which it was derived.

Earlier data (13, 14) and our recent work (unpublished data) indicate that the *SURL* element family consists of subfamilies. Each subfamily contains from a few to 50 members (13). The subfamilies have been distinguished by the DNA sequence divergence of a conserved segment of the RT gene coding region. Members of each subfamily have closely similar sequences (1–10% divergent), whereas this DNA sequence region of each subfamily is greatly divergent from that of other subfamilies (30–45%). Interestingly, examples of intermediate divergence (10–30%) are rare or nonexistent. The implication is that copies of the RL elements that have been inserted into the genome drift in sequence, and the observation of stop codons in required coding regions suggests that many pseudogenes or "pseudocopies" are present. A similar pattern (termed superfamily and families) has been observed for the *Ta* elements of *Arabidopsis* (10) that are similar to *copia* of *Drosophila* (15). However, these elements are less similar to retroviruses than are the *SURL* elements. The means by which sea urchin RL elements were first identified and cloned will be described elsewhere, as will be the data supporting the subfamily structure, while this article focuses on their major characteristics as implied by DNA sequences.

MATERIALS AND METHODS

Sea urchin DNAs were extracted from sperm or eggs, digested with *Eco*RI, and cloned into Lambda ZAP II (Stratagene). Libraries were screened by standard procedures (16) with nick-translated *Hin*dIII fragments (1.3 and 2.0 kb) from clone SURL1-21 (Sp) that span much of the *gag* and *pol* genes. Positive clones were subcloned into pBluescript SK(−) by using the *in vivo* excision procedure (Stratagene). DNA was sequenced by the dideoxynucleotide chain-termination method using Sequenase (17). Subclones for sequencing were obtained by using the exonuclease III/mung bean nuclease procedure (16) and subcloning *Sau*3A restriction fragments into the *Bam*HI site of pBluescript SK(−). Extension of synthetic primers made from known sequences was used to fill in missing sequences and to obtain sequences for the RT gene coding region from clones for different species. Ninety-five percent of the sequences were obtained in both directions; those that were not were sequenced multiple times in the same direction. Alignments were made using the FASTA and PAPA programs (18, 19). Comparisons of synonymous versus replacement substitutions were made by using a program supplied by W.-H. Li (20).

RESULTS

Sequence and LTRs. The example that has been fully sequenced† [SURL1-3 (Tg)] from *T. gratilla* is 5266 base pairs (bp) long, has a 5-bp target site duplication (CCACC) of host DNA, and has identical 254-bp LTRs. The LTRs (Fig. 1) contain TG . . . . CA terminal inverted repeats, a putative

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: LTR, long terminal repeat; RT, reverse transcriptase; ORF, open reading frame; RL elements, retroviral-like elements; *SURL* elements, sea urchin RL elements.

*Present address: Department of Biology, University of California, Riverside, CA 92521.

†The sequence reported in this paper has been deposited in the GenBank data base (accession no. M75723).

```
TGTAGTGATAGCTATACTTATATAGTCTGCAAACACAGACTAGGTAGACTGA
a

CCTGACATTTATTACCTATGACTCATGCCCTAGGTCAGTATACTATAAAC
                                              b

AATGATGCTCTTGTGTAATTGAGCACATGGTTCAAAGATGTCGACTGAGT

AGTACCTAATAAAGATTTCCTATATTATAAGTTCAACTAAAGCTCTCCGA
        c                        d

CTGGTCTTTTATCCATTATACTGAAGTCAGGATTTATAATGGAACCAAGCCA
                                                    a
```

FIG. 1. Nucleotide sequence of the LTR of clone SURL1-3 (Tg). a, Terminal inverted repeats; b, putative TATA box; c, putative polyadenylylation signal; d, putative polyadenylylation site.

promoter region, a polyadenylylation signal, and a polyadenylylation site. Immediately interior to the 5′ LTR is a putative tRNA primer binding site for (−)-strand synthesis that begins with the sequence TGG. Sea urchin nuclear tRNAs have not been characterized; however, starting with the TGG sequence, SURL1-3 (Tg) shares 11 of 12 bases with the methionine tRNA primer binding site of both RL elements *Tal-3* of *Arabidopsis* (10) and *del* of lily (9). *Ty1* (3) and *Ty3* (4) of yeast are also both inferred to be primed with methionine tRNA. Interior to the 3′ LTR, in turn, is a 13-bp polypurine region (AAAGAAAGGGAGG), a characteristic of (+)-strand priming sites in retroviruses (21). Such polypurine tracts are also present in *Ty3* (4) and *Tf* elements (5) of yeast. Immediately upstream of this is a string of eight pyrimidines, a feature shared with certain retrotransposons (22, 23) and plant pararetroviruses (24, 25). The internal domain contains a single 3927-bp open reading frame (ORF), and the predicted 1309 amino acid sequence includes both the *gag* and *pol* regions. A single ORF including both the *gag* and *pol* genes is unusual but is known in the *Tf* elements (5), in *copia* and its close relatives (10), and in *micropia* (26). The *pol* region encodes protease, RT, RNase H, and integrase functions in the same order as in most retroviruses, based on amino acid sequence similarities of which examples are shown in Figs. 2, 3, and 4.

```
A                    *  *     *     *        *  *     *     *
SURL1-21(Sp) DACSRCGNR-HSKDKCPA-QDQTCKKC-KQNHFAKCCKT
SURL1-3 (Tg) KTCSRCGNR-HSKDKCPA-QGKTCKKCGKQNHFAKCCRT
HIV          VKCFNCGKEGHTARNCRAPRKKGCWKCGKEGHQMKDCTE
IAPm         RTCFNCGKPGHFKKDCRAPDKQ>CSKCGKGYHRADQCRS

B                     *  *
SURL1-21 (Sp)    VDFKVDTGAQANLLP
SURL1-3  (Tg)    VEFKIDTGAQANLLP
MLV              VTFLVDTGAQHSVLT
CAMV             LHCFVDTGASLCIAS
IAPm             FEGILDTGADKSIIS
17.6             LKCLIDTGSTVNMTS
RSV              ITALLDSGADITIIS
Copia            CGFVLDSGASDHLIN
Ty3              VKTLFDSGSPTSFIR
```

FIG. 2. (*A*) Multiple alignment of proteins in single-letter code showing the RNA binding sites. Two copies of the RNA binding site are part of the nucleocapsid protein of many retroviruses, but known retrotransposons encode only one copy or none. The spacing between the two copies varies, being six amino acids in SURL1, 7 in human immunodeficiency virus (HIV), and 11 in mouse intracisternal A particle (IAPm). SURL1-21 (Sp) is from *S. purpuratus* and SURL1-3 (Tg) is from *T. gratilla*. Asterisks above the top sequence show the presumptive zinc-finger pattern of cysteines and histidines; bold symbols show recognizable conserved amino acid positions. (*B*) Multiple alignment showing the active site of the putative protease gene. Asterisks mark the apparently invariable amino acids; bold letters show all matches with the SURL sequences. MLV, mouse leukemia retrovirus; CaMV, cauliflower mosaic pararetrovirus; RSV, Rous sarcoma virus. *17.6* and *copia* are retrotransposons from *Drosophila*, and *Ty3* is a retrotransposon from yeast.

Relationships of the *SURL* Elements. Amino acid sequence comparison with other RL elements (Fig. 2*A*) shows two RNA binding sites that are presumably part of a nucleocapsid

```
                    10         20         30         40         50         60
SURL6 (Et) VNtmMDDIIV wGstqaEHdt rLkqVLsiaR KsNLKLnREK CEFnVnKlTY iGdLISEneV
SURL5 (Lv)         vV wGstqEEHdd rLRkVLdiaq KsNLKLnRDK CEFnVnQmTF iGdLISqDGV
SURL1 (Sf) VNtmMDDVIV wVTtqqEHdd rLREVLsivR KmNLKLnKDK CEFnVKKlTF iGdLISDQrV
SURL1 (Sp) VNtmMDDVIV wrTtq*Erdd rLRkVLsivR KmNLKLnKDK CEFnVKKlTF iGdLIIDQGf
SURL1 (Tg) VNtmMDDVIV wGTtqqEHdN rLREVLsiaR rmNLKLnKDK CEFsVKKlTF iGdLISDQGV
Yeast Ty3  VNvylDDIli fsespEEHwk hLdtVLErlK neNLivkKkK CkFaseeteF lGysIgiQki
Dm 17.6    clvylDDIIV fsTsldEHlq sLglVfEkla KaNLKLqlDK CEFlkqetTF lGhvltpDGi
Dm 297     clvylDDIIi fsTsltEHlN siqlVftkla daNLKLqlDK CEFlkKeanF lGhivtpDGi
Dm 412     aflyMDDlIV iGcsekhmlk nLtEVfgkcR eyNLKLhpEK CsFfmhevTF lGhkctDkGi
Dm gypsy   cyvyvDDVIi fsenesdHvr hidtVLkcli daNmrvsqEK trFfkesvey lGfivSkDGt
Carn. ring ccvyvDDIlV fsTgrkEHyi hvlniLrrce KlgiiLsKkK aqlfkeKinF lG-LeIDQGt
Caul. mosa ccvyvDDIlV fsnneEdHll hvamiLqkcn qhgiiLsKkK aqlfkKKinF lG-LeIDeGt
Pine IFG7  VlvffDDIli ynkswkdHve hvdrVLqlle ekkLyakRsK CfFvVqevey lGhivSleGV
AKV murine llqyvDDIll aaTseldcqq gtRalLltlg nlgyrasakK aqlcqKQvky lGyLlkEgqr
MOMLV      llqyvDDlll aaTseldcqq gtRalLqtlg nlgyrasakK aqicqKQvky lGyLlkEgqr
                   70         80         90        100
SURL6 (Et)   qPDPqKVAAI RNmeRPKSKq dVQRFLGMvN YqgKFIPDLA       73
SURL5 (Lv)   RPDPKKVAAI RNmaRPTCKq dIQRFLGMIN YqaKFIPNLS       69
SURL1 (Sf)   qtDPKKVAAI lNmeRPKyre nVQRFLGMIN YqgKFIPDLS       86
SURL1 (Sp)   qPDPKKVsAI lNieRPKCrK dVQRFLGMnN YqgKFIPDLS    % Identity
SURL1 (Tg)   qPDPKKVsAI lNmeRPKCrK dVQRFLGMIN YqgKFIPDLS       88
Yeast TY3    aPlqhKcAAI RdfptPKtvK qaQRFLGMIN YyrrFIPNcS       37
Dm 17.6      kPnPeKieAI qkypiPTkpK eIkaFLGltg YyrKFIPNfA       35
Dm 297       kPnPiKVkAI vsypiPTkdK eIraFLGltg YyrKFIPNyA       34
Dm 412       lPDdKKydvI qNypvPhdad sarRFvafcN YyrrFIkNfA       32
Dm gypsy     ksDPeKVkAI qeypePdCvy kVrsFLGlas YyrvFIkDfA       27
Carn. ring   hPqnhilehI hkfdRiedKK qlQRFLGilt YasdyIPkLA       29
Caul. mosa   hkqghilehI nkfdtledKK qlQRFLGilt YasdyIPkLA       26
Pine IFG7    kvyPnKikAI kewkiPTSiK hlrgFLGltg Y-rKFakNyg       26
AKV murine   wltearketv mgqptPKtpr qlreFLGtag fcrlwIPgfA       16
MOMLV        Rltearkety mgqptPKtpr qlreFLGtag fcrlwIPgfA       16
```

FIG. 3. Alignment through the conserved region near the universal Asp-Asp of RT amino acid sequences. The listed SURL amino acid sequences are translations of recently determined DNA sequences. No insertions or deletions have been used in the alignment for the upper group including gypsy. Below that only a half dozen single nucleotide deletions have been used altogether. Amino acids shown in boldface letters are shared among the products of a *SURL* element and four or more others; where there is agreement with the translation of a *SURL* element, both are capitalized. The asterisk indicates a stop codon. The percent identity is relative to SURL1-21 (Sp) clone.

```
Sp  V_Q_S C_T_S C_L M Y R P K Q Q_A E S L H P_H D V   P_N R P W_E K V A V
Tg  V_Q_S C_T_S C_L M Y K P K Q Q_A E S L H P_H A V   P_S R P W_E K V A V
HU  V_Q_H C_T_Q C_Q V L H L P T Q_E A G V N P_R G L C P_N A L W_Q M D V T

Sp  Y L F T L N N R E Y_M   V_I V_D_Y Y_S_Q F_I_E V C T M T_S T S_S K A V_
Tg  D L F T L N K R E Y_M   V_I V_D_Y Y_S_Q F_I_E V C T M T_S T S_S K A V_
HU  H V P S F G R L S Y_V H V_T V_D_T Y_S_H F_I_W A T C Q T_G E S_T S H V_

Sp  I N H_M K A V F_A_R H G_T P_C E L M S D_N_G_P_Q F A S_Q E F_Q S F_V K
Tg  I N H_M K A I F_A_R H G_T P_C E L M S D_N_G_P_Q F A S_Q E F_K S F_A K
HU  K K H_L L S C F_A_V M G_V P_E K I K T D_N_G_P_G Y C S_K A F_Q K F_L S

Sp  E W_D F H H_T_T_S S P_Y Y P Q_S N G L A E_N A V K I V K_K
Tg  E W_D F H H_T_T_S S P_H Y P Q_S N G L A E_N A V K I V K_K
HU  Q W_K I S H_T_T_G I P_Y N S Q_G Q A I V E_R T N R T L K_T
```

FIG. 4. Integrase gene alignment between human endogenous retrovirus HERV-K10 (HU) and clones SURL1-21 (Sp) and SURL1-3 (Tg) as shown by the corresponding amino acid sequences in single-letter code. The percent amino acid identity with the human retrovirus is 32.6% for SURL1-21 (Sp) and 30.3% for SURL1-3 (Tg). Other retroviruses are less similar in this region, some having as low as 19% amino acid sequence identity. Alignment was done with PAPA programs kindly supplied by R. F. Doolittle. Amino acid symbol is followed by a lowered dash if all three sequences match.

gene. Thus, in common with most retroviruses (27), both *SURL* elements contain two RNA binding sites, each possessing a presumptive zinc-finger pattern of cysteines and histidines (28). Previously reported retrotransposons either lack RNA binding sites, as is common among *Drosophila* retrotransposons related to *gypsy*, or contain a single site as in *copia*, *del*, and *Ty3* elements (4, 9, 15). A short region of the presumptive protease gene is shown in Fig. 2*B*. The conserved pattern of Asp-(Ser or Thr)-Gly, which is part of the active site of the enzyme that cleaves the *gag-pol* polyprotein, is held in common among a wide range of retroviruses and retrotransposons as well as many other proteases (4). Sequence comparisons show that long regions of the *pol* gene are conserved, and thus they may be useful indicators of RL element relationships (18, 29). Fig. 3 shows a representative 100-amino acid region of the RT gene, and this comparison suggests that the *SURL* elements are more closely related to a branch of retrotransposons that includes *gypsy* of *Drosophila* (5, 18, 29), *Ty3* of yeast (4), and plant pararetroviruses (24, 25) than they are to representative retroviruses. In contrast to this observation, a portion of the integrase gene is more similar in amino acid sequence to certain retroviruses than to retrotransposons (Fig. 4), with the exception of the *Mag* retrotransposon from the silkmoth (29). A question is whether the *SURL* elements can be classed as LTR-containing retrotransposons or retroviruses. The RT amino acid sequence relationships suggest that they are retrotransposons. Furthermore, in retroviruses the *env* gene is in a separate ORF just 3′ of the integrase gene, while in the SURL1 (Tg) clone there is no such separate ORF that appears to be a candidate for an *env* gene. The nucleocapsid and integrase amino acid sequences suggest retroviral affinities, but these enzymes are not concerned with the infectious extracellular phase of retroviruses. Less conventional infectious processes could occur among retrotransposons; for example, the lepidopteran retrotransposon *TED* may be capable of expanding its host range via insertion in a baculovirus genome that shuttles it to a new host (7). "Helper" interactions may exist with other viruses and a new domain such as *SURL* elements may contain surprises, especially in view of evidence that viruses are abundant in sea water (30). However, up to this time no retroviruses have been identified among marine viruses.

History of Active *SURL* Elements. The presence of stop codons and divergence within the subfamily suggest that most *SURL1* subfamily members are inactive inserted copies that have been subject to drift. However, comparison of DNA and amino acid sequence divergence among elements in distant sea urchin species indicates that there are also active copies of *SURL* elements. The comparison of clones SURL1-3 (Tg) and

Table 1. Synonymous and replacement substitution differences per site between ORFs of *SURL* elements

Clones compared	ORF	Substitution difference per site*		Ratio	Approximate gene region
		Synonymous†	Replacement		
Sp/Tg	Segment‡				
	1	0.37 (0.07)	0.04 (0.010)	8.7	*gag*
	2	0.68 (0.11)	0.10 (0.016)	6.6	*gag* + protease
	3	0.49 (0.08)	0.04 (0.01)	11.5	RT
	4	0.62 (0.10)	0.05 (0.01)	12.9	RT + RNase H
	5	0.72 (0.11)	0.05 (0.01)	13.3	RHase H + Integrase§
	6	0.56 (0.09)	0.05 (0.01)	11.3	Integrase + ?
	7	0.40 (0.10)	0.1 (0.02)	4.1	?
	Whole	0.56	0.060	9.3	
Sp/Sf	700-nt¶	0.14 (0.03)	0.04 (0.01)	3.2	RT segment
Tg/Sf		0.42 (0.07)	0.05 (0.01)	8.6	RT segment
Sp/Et		3.12 (1.3)	0.17 (0.02)	18.4	RT segment
Sf/Et		2.95 (0.96)	0.15 (0.02)	19.1	RT segment

Sp, SURL1-21 (Sp); Tg, SURL1-3 (Tg); Sf, SURL1-7 (Sf); Et, SURL6-1 (Et).
*Standard errors are shown in parentheses.
†Calculated with the LWL85 program kindly supplied by Wen-Hsiung Li.
‡The ORF is divided into 200-amino acid sections except for segment 7, which has only 105 amino acids.
§The integrase gene probably starts about amino acid 895 and may not extend the full 411 amino acids to the first stop codon; this would leave room only for a very short *env* gene if it exists at all.
¶Comparisons of a 700-nucleotide (nt) region of the RT gene.

SURL1-21 (Sp) (see Table 1) shows a great excess of synonymous substitutions versus those causing amino acid replacement (per site) over the entire ORF including *gag* and *pol* genes. This shows that these coding regions have evolved under sequence-dependent selection. In addition a 702-bp region of the RT and RNase H genes was sequenced for clones derived from a *SURL1* element of *S. franciscanus* DNA and a member of another *SURL* subfamily from the very distant sea urchin *E. tribuloides*. Table 1 also shows that for this region there is an excess of synonymous versus nonsynonymous substitutions in all comparisons, providing further support for the concept that the RT protein evolved under strong selection over much of the evolutionary history of echinoids. Estimates for the lower bounds on divergence times between *S. purpuratus* and *S. franciscanus*, *T. gratilla*, and *E. tribuloides*, respectively, are 20, 45, and 250 million years (31, 32). These estimates give rates of synonymous substitution of 0.35%, 0.49%, and 0.62% divergence per lineage per million years. These rates are similar to the rate of 0.5% divergence per lineage per million years calculated from single-copy DNA hybridization studies (33). Thus, it appears that synonymous rates of change among RT genes of sea urchin retrotransposons are about equal to the single-copy drift rate, which is a very slow rate compared with that of the known vertebrate retroviruses. We may conclude from the freely occurring synonymous substitutions and the selection against amino acid replacements that active copies, dependent on these genes for replication, have been the main line of inheritance of *SURL* elements during echinoid evolution.

It is interesting that synonymous/nonsynonymous ratios are higher for comparisons between more distantly related taxa. Thus, the ratio is about 3 for comparisons between *S. purpuratus* and *S. franciscanus*, 8.6–12.9 for comparisons between *Tripneustes* and *Strongylocentrotus*, and 18.4–19.1 for comparisons between euechinoids and cidaroids. In part this may be due to the toleration by these elements of a number of amino acid replacements in weakly selected positions. Other replacements may be more strongly selected; as a result, rates of nonsynonymous divergence decelerate, as recently noted in plant nuclear genes (34). In addition, individual elements in a genome at any given time evolve as pseudogenes and accumulate nonsynonymous changes more rapidly than synonymous changes, affecting the comparisons between elements in closely related species. However long-term survival of retrotransposons requires functional copies that evolve under strong sequence-dependent selection, and distant comparisons are probably less affected by the presence of pseudogenes in the copies we have examined.

DISCUSSION

The sequence evidence combined with earlier work shows that sea urchins include in their DNA many interspersed copies of RL elements. They have been present in each sea urchin species we have examined, and, as described in the introduction, they are being observed in an increasing number of species. There are no reports known to us of the total absence of such elements in any eukaryotic genome, although this would be difficult to prove with the limitations of presently existing probes. It is not unlikely that they are a generally occurring feature of eukaryotic DNA. The potential common occurrence and the evidence described above of active copies evolving under sequence-dependent selection bear on the issue of the fluidity of eukaryotic genomes. An element that was once considered simply a repeated sequence (13, 14) is shown to have had an active replication and DNA insertion cycle in the sea urchin genome for many millions of years. Mobile elements that insert into the genome are a cause of evolutionary variation because of disruption of gene regions and because of the influence of enhancers and other control sequences of the inserted element on adjacent genes. Most such events are probably damaging and are selected against but occasionally may have positive value as in the case of a mouse provirus insertion that usefully imposed androgen regulation on an adjacent gene (35). The SURL elements are probably a significant part of the panoply of mobile elements that give rise to repeated sequences and cause important evolutionary variation.

We thank R. F. Doolittle and Wen-Hsiung Li for programs and are grateful for a *T. gratilla* phage λ library prepared by Gregory Dolecki at Kewalo Marine Laboratory, Hawaii. This work was supported by the Lucille P. Markey Charitable Trust and the National Institutes of Health.

1. Temin, H. M. (1985) *Mol. Biol. Evol.* **2**, 455–468.
2. McDonald, J. F. (1990) *BioScience* **40**, 183–191.
3. Fink, G. R., Boeke, J. D. & Garfinkel, D. J. (1986) *Trends Genet.* **2**, 118–123.
4. Hansen, L. J., Chalker, D. L. & Sandmeyer, S. B. (1988) *Mol. Cell. Biol.* **8**, 5245–5256.
5. Levin, H. L., Weaver, D. C. & Boeke, J. D. (1990) *Mol. Cell. Biol.* **10**, 6791–6798.
6. Bingham, P. M. & Zachar, Z. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 485–502.
7. Friesen, P. D. & Nissen, M. S. (1990) *Mol. Cell. Biol.* **10**, 3067–3077.
8. Kossack, D. S. (1989) Ph.D. dissertation (Univ. of California, Davis).
9. Smyth, D. R., Kalitsis, P., Joseph, J. L. & Sentry, J. W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 5015–5019.
10. Konieczny, A., Voytas, D. F., Cummings, M. P. & Ausubel, F. M. (1991) *Genetics* **127**, 801–809.
11. Grandbastien, M.-A., Spielmann, A. & Caboche, M. (1989) *Nature (London)* **337**, 376–380.
12. Schiff, R., Itin, A. & Keshet, E. (1991) *Genes Dev.* **5**, 520–532.
13. Scheller, R. H., Anderson, D. M., Posakony, J. W., McAllister, L. B., Britten, R. J. & Davidson, E. H. (1981) *J. Mol. Biol.* **149**, 15–39.
14. Anderson, D. M., Scheller, R. H., Posakony, J. W., McAllister, L. B., Trabert, S. G., Beall, C., Britten, R. J. & Davidson, E. H. (1981) *J. Mol. Biol.* **145**, 5–28.
15. Mount, S. M. & Rubin, G. M. (1985) *Mol. Cell. Biol.* **5**, 1630–1638.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Doolittle, R. F., Feng, D.-F., Johnson, M. S. & McClure, M. A. (1989) *Q. Rev. Biol.* **64**, 1–30.
19. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2444–2448.
20. Li, W.-H., Wu, C.-I. & Luo, C.-C. (1985) *Mol. Biol. Evol.* **2**, 150–174.
21. Varmus, H. & Brown, P. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 53–108.
22. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. & Yuki, S. (1984) *Nature (London)* **312**, 659–661.
23. Inouye, S., Yuki, S. & Saigo, K. (1986) *Eur. J. Biochem.* **154**, 417–425.
24. Franck, A., Guilley, H., Jonard, G., Richards, K. & Hirth, L. (1980) *Cell* **21**, 285–294.
25. Hull, R., Sadler, J. & Longstaff, M. (1986) *EMBO J.* **5**, 3083–3090.
26. Lankenau, D.-H., Huijser, P., Jansen, E., Miedema, K. & Hennig, W. (1988) *J. Mol. Biol.* **204**, 233–246.
27. Covey, S. N. (1986) *Nucleic Acids Res.* **14**, 623–633.
28. Green, L. M. & Berg, J. M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4047–4051.
29. Xiong, Y. & Eickbush, T. H. (1990) *EMBO J.* **9**, 3353–3362.
30. Suttle, C. A., Chan, A. M. & Cottrell, M. T. (1990) *Nature (London)* **347**, 467–469.
31. Smith, A. B. (1988) *Mol. Biol. Evol.* **5**, 345–365.
32. Smith, A. B. (1989) *Cladistics* **5**, 321–344.
33. Britten, R. J. (1986) *Science* **231**, 1393–1398.
34. Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) *J. Mol. Evol.* **29**, 208–211.
35. Stavenhagen, J. B. & Robins, D. M. (1988) *Cell* **55**, 247–254.