SHORT COMMUNICATION

CHUNG-MO PARK, JOHN D. LOPINSKI, JENNY MASUDA, TZY-HWA TZENG, and JEREMY A. BRUENN[1]

Department of Biological Sciences, State University of New York at Buffalo, Buffalo, New York 14260

Received October 13, 1995; accepted December 22, 1995

Two double-stranded RNA viruses exist as permanent persistent infections of the yeast *Saccharomyces cerevisiae:* ScVL1 and ScVLa. Both belong to the Totiviridae, which include a number of fungal and protozoan double-stranded RNA viruses. Although ScVL1 and ScVLa share the same genomic organization and mode of expression and coexist in the same cells, they show no evidence of recombination: with one limited exception, sequence conservation is detectable only in regions conserved in all totiviruses. Both have two open reading frames on their single essential RNAs: *cap* (encoding a capsid polypeptide) and *pol* (encoding an RNA-dependent RNA polymerase). The ScVLa virus, like ScVL1, appears to express its Pol domain by a −1 translational frameshift.   © 1996 Academic Press, Inc.

We have shown that a group of double-stranded RNA (dsRNA) viruses of lower eucaryotes are more closely related to each other than they are to any other viruses (*1*). These are members of the Totiviridae, which have a single essential dsRNA segment; most are noninfectious. At least in the fungi, it appears to be quite common for a given species to be a carrier of several such viruses simultaneously. This was first conclusively demonstrated for the yeast *Saccharomyces cerevisiae,* most laboratory strains of which have at least two such viruses (*2*). We originally named their genomic RNAs L1 and La (*2*). In an average laboratory strain of yeast, the *S. cerevisiae* La virus (ScVLa) is present at about 5% the level of the L1 virus (ScVL1) (*3*). La is also known as L-BC and L1 as L-A (*4*).

The entire sequence of the essential segment of the ScVL1 virus (L1) has been known for some time (*5, 6*). We have cloned and sequenced cDNAs representing the entire sequence of La, which has a genomic organization similar to that of L1. La dsRNA was extracted from strain 299 (*MAT a lys1 mak3-1* [*KIL-O*] *KIL-R⁺ La*) from the Yeast Genetics Stock Center. dsRNA was isolated by phenol extraction from cells disrupted by glass beads, followed by CF11 chromatography (*7*). The original La cDNA clone was derived by the RNase H method (*8*) by Martin Nemeroff. Subsequent clones were made by the same method, using primers derived from the known sequence for the first strand. Sequencing was by the dideoxy method (*9*). The set of six cDNA clones used for sequencing La and the sequencing runs performed, as well as the genomic organization of La, are outlined in Fig. 1. Sequencing of

both strands is complete, except for the very small regions at the ends which were determined by RNA sequencing. All overlapping regions of independently isolated clones were identical in sequence. The entire sequence of the viral plus strand of 4615 bases (given as DNA here) is shown in Fig. 2.

The 3′ end sequences of both strands of total L dsRNA were previously determined by ^{32}pCp labeling and direct RNA sequencing (*10*). It is now clear that the 3′ T1 oligonucleotides of the plus strands of both L1 and La are (G)CA, while the 3′ T1 oligonucleotides of the minus strands are (G)AAAAAUUUUUCA (L1) and (G)AAAAAU-UCA (La). This last is the minor oligonucleotide 6 of the T1 digest of ^{32}pCp-labeled total L dsRNA (*10*). Both L1 and La therefore have 3′ terminal A residues not encoded in the template strands.

The genomic organization of La is identical to that of
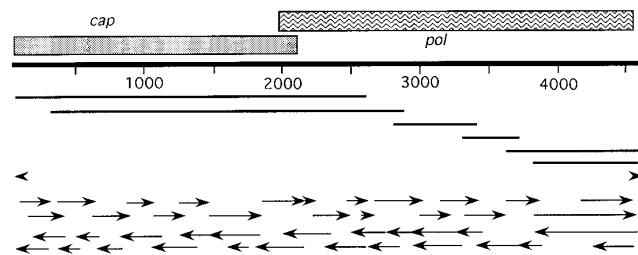
FIG. 1. Cloning and sequencing strategy for La. The two open reading frames are labeled above a scale (in bases on the plus strand). Below the scale the individual cDNA clones sequenced are indicated as straight lines. The orientation of sequencing runs is indicated by arrows. The two short RNA sequencing runs are above and the DNA sequencing runs below. All regions (with the exception of short regions at the 5′ and 3′ ends of the plus strand) were sequenced on both strands, usually with multiple runs.

[1] To whom reprint requests should be addressed.

```
1    GAATTTTTCGGTGAACCGGAATTATGTCGTCTCTGTTAAATTCATTACTACCAGAATATT   60
                   M  S  S  L  L  N  S  L  L  P  E  Y  F           13
61   TTAAACCTAAAACTAATTTGAATATCAACTCTTCTAGGGTCCAATATGGCTTTAATGCTC  120
14    K  P  K  T  N  L  N  I  N  S  S  R  V  Q  Y  G  F  N  A  R    33
121  GCATTGATATGCAGTATGAAGACGATAGTGGGACTAGAAAAGGCTCAAGACCCAATGCAT  180
34    I  D  M  Q  Y  E  D  D  S  G  T  R  K  G  S  R  P  N  A  R    53
181  TTATGTCTAACACAGTTGCTTTTATAGGAAACTATGAAGGTATTATTGTTGATGACATTC  240
54    M  S  N  T  V  A  F  I  G  N  Y  E  G  I  I  V  D  D  I  P    73
241  CGATATTGGATGGTCTTAGGGCCGACATTTTTGATACTCATGGTGACTTAGACATGGGCC  300
74    I  L  D  G  L  R  A  D  I  F  D  T  H  G  D  L  D  M  G  L    93
301  TCGTTGAGGATGCATTGTCTAAGAGTACCATGATTAGAAGGAATGTACCAACTTATACTG  360
94    V  E  D  A  L  S  K  S  T  M  I  R  R  N  V  P  T  Y  T  A    113
361  CTTACGCGAGTGAATTACTGTATAAGAGAAACCTTACATCTCTATTTTACAATATGCTCC  420
114   Y  A  S  E  L  L  Y  K  R  N  L  T  S  L  F  Y  N  M  L  R    133
421  GTTTATACTACATTAAAAAAATGGGGCAGTATTAAGTATGAAAAAGATGCCATCTTTTATG 480
134   L  Y  Y  I  K  K  W  G  S  I  K  Y  E  K  D  A  I  F  Y  D    153
481  ATAATGGCCACGCCTGTCTTTTAAACAGGCAATTGTTTCCAAAGTCTCGTGATGCTCTT   540
154   N  G  H  A  C  L  N  R  Q  L  F  P  K  S  R  D  A  S  L       173
541  TGGAATCAAGCCTCTCTTTGCCTGAGGCTGAAATTGCAATGCGTTGATCCTGGCCTGGAAT 600
174   E  S  S  L  S  L  P  E  A  E  I  A  M  L  D  P  G  L  E  F    193
601  TTCCAGAAGAGGATGTGCCTGCAATTTTATGGCACGGCAGAGTGTCATCCAGAGCAACGT  660
194   P  E  E  D  V  P  A  I  L  W  H  G  R  V  S  S  R  A  T  C    213
661  GTATCTTAGGGCAAGCTTGCTCAGAGTTCGCGCCTCTGGCCCCCCTTTTCGATTGCGCATT 720
214   I  L  G  Q  A  C  S  E  F  A  P  L  A  P  F  S  I  A  H  Y    233
721  ATTCACCACAATTGACGAGAAAACTATTTGTCAATGCGCCCGCTGGGATTGAGCCTAGCT  780
234   S  P  Q  L  T  R  K  L  F  V  N  A  P  A  G  I  E  P  S  S    253
781  CCGGGCGGTATACTCACGAGGATGTAAAAGATGCGATTACGATCCTTGTGTCTGCAAACC  840
254   G  R  Y  T  H  E  D  V  K  D  A  I  T  I  L  V  S  A  N  Q    273
841  AGGCTTATACTGACTTTGAAGCAGCATACTTGATGCTTGCTCAAACGTTGGTCTCACCTG  900
274   A  Y  T  D  F  E  A  A  Y  L  M  L  A  Q  T  L  V  S  P  V    293
901  TACCACGCACTGCCGAAGCAAGTGCATGGTTCATCAATGCTGGCATGGTCAATATGCCAA  960
294   P  R  T  A  E  A  S  A  W  F  I  N  A  G  M  V  N  M  P  T    313
961  CTTTGTCATGTGCAAATGGTTATTATCCAGCACTGACCAATGTCAATCCTTACCACCGAT 1020
314   L  S  C  A  N  G  Y  Y  P  A  L  T  N  V  N  P  Y  H  R  L    333
1021 TAGACACATGGAAAGATACGTTAAATCATTGGGTGGCTATTCCCGACATGCTGTTTTACC 1080
334   D  T  W  K  D  T  L  N  H  W  V  A  Y  P  D  M  L  F  Y  H    353
1081 ATTCAGTGGCAATGATTGAGAGCTGCTATGTTGAACTCGGGAATGTGGCTCGTGTGTCAG 1140
354   S  V  A  M  I  E  S  C  Y  V  E  L  G  N  V  A  R  V  S  D    373
1141 ACAGTGATGCAATAAACAAATACACTTTCACTGAGCTATCAGTGCAAGGACGGCCTGTTA 1200
374   S  D  A  I  N  K  Y  T  F  T  E  L  S  V  Q  G  R  P  V  M    393
1201 TGAATCGAGGAATTATTGTAGATCTGACATCTGTGGCAATGCGTACTGGAGGGAGATCT  1260
394   N  R  G  I  I  V  D  L  T  L  V  A  M  R  T  G  R  E  I  S    413
1261 CACTACCTTACCCGGTCAGCTGTGGCCTGACCGTACAGACGCGGTTATTGCAAGGTACTG 1320
414   L  P  Y  P  V  S  C  G  L  T  R  T  D  A  L  L  Q  G  T  E    433
1321 AGATTCACGTTCCAGTTGTTGTCAAAGATATCGACATGCCCCAGTATTACAACGCGATTG 1380
434   I  H  V  P  V  V  K  D  I  D  M  P  Q  Y  Y  N  A  I  D       453
1381 ATAAGGATGTTATTGAGGGGCAGGAAACTGTGATTAAAGTGAAACATGCTGCCACCAGCTA 1440
454   K  D  V  I  E  G  Q  E  T  V  I  K  V  K  Q  L  P  P  A  M    473
1441 TGTATCCAATTTATACTTACGGGATCAACACTACTGAATTCTATTCTGACCATTTTGAAG 1500
474   Y  P  I  Y  T  Y  G  I  N  T  T  E  F  Y  S  D  H  F  E  D    493
1501 ACCAGGTACAAGTTGAAATGGCACCAATCGATAATGGAAAAGCAGTTTTTAACGATGCAA 1560
494   Q  V  Q  V  E  M  A  P  I  D  N  G  K  A  V  F  N  D  A  R    513
1561 GAAAGTTTTCGAAATTTATGTCCATAATGCGCATGATGGGGAATGATGTTACTGCTACTG 1620
514   K  F  S  K  F  M  S  I  M  R  M  M  G  N  D  V  T  A  T  D    533
1621 ATTTAGTTACAGGTAGAAAAGTGTCGAATTGGGCCGACAACTCATCAGGGCGTTTCTTGT 1680
534   L  V  T  G  R  K  V  S  N  W  A  D  N  S  S  G  R  F  L  Y    553

1681 ACACGGATGTGAAGTATGAAGGACAAACTGCCTTTTTGGTTGATATGGATACTGTCAAGG 1740
554   T  D  V  K  Y  E  G  Q  T  A  F  L  V  D  M  D  T  V  K  A    573
1741 CGGAGAGACCACTGTTGGGTGTCAATTGTTGATCCTAATGGTACAATGAACTTGTCATATA 1800
574   R  D  H  C  W  V  S  I  V  D  P  N  G  T  M  N  L  S  Y  K    593
1801 AGATGACCAATTTTAGAGCAGCAATGTTTTCTAGAAACAAGCCCTTGTATATGACAGGGG 1860
594   M  T  N  F  R  A  A  M  F  S  R  N  K  P  L  Y  M  T  G  G    613
1861 GGTCAGTCAGGACCATAGCTACTGGCAATTATCGAGATGCTGCTGAAAGATTACGTGCAA 1920
614   S  V  R  T  I  A  T  G  N  Y  R  D  A  A  E  R  L  R  A  M    633
1921 TGGATGAAACGCTCAGATTAAAACCTTTTAAGATTACTGAGAAGTTGGATTTTCGTGTAG 1980
634   D  E  T  L  R  L  K  P  F  K  I  T  E  K  L  D  F  R  V  A    653
649                                           G  F  S  C  S        653
1981 CAGCTTACGCGATACCAAGTTTGTCGGGCAGCAATATGCCATCCTTACACCATCAGGAAC 2040
654   A  Y  A  I  P  S  L  S  G  S  N  M  P  S  L  H  H  Q  E  Q    673
2041 AACTACAGATATCAGAAGTGGACGCGGAACCAATCAATCCTATAGGAGAGGACGAACTTC 2100
674   L  Q  I  S  E  V  D  A  E  P  I  N  P  I  G  E  D  E  L  P    693
2101 CACCGGATATAGAATAGGTGTCGAAGACGATGAGGACTTAGATATTGGTACGGTCAAATA 2160
694   P  D  I  E                                                   697
694   T  G  Y  R  I  G  V  E  D  D  E  D  L  D  I  G  T  V  K  Y    713
2161 CATTGTGCCATTGTATTGAACGGTGATAATGTGGCACAAAATTGTTTAGAAGCAACACA  2220
714   I  V  P  L  Y  L  N  G  D  N  V  A  Q  N  C  L  E  A  T  H    733
2221 CGTGCTTATCAAAGCTTGTAGTATTGCGAACCGGATTGTAGATGACGGAGAGGGTCACTG 2280
734   V  L  I  K  A  C  S  I  A  N  R  I  V  D  D  G  E  G  H  C    753
2281 TTTCACACACAGCAAGGGCTGGCGCAGCAGTGGATCTTCCATAGGGGGGAGATGATATTTGT 2340
754   F  T  Q  Q  G  L  A  Q  Q  W  I  F  H  R  G  E  M  I  F  V    773
2341 GAAGGCGGGTACGCATTGGTCAACTCAATGCATATTATGTAGACTATAAGAACGTCACAAA 2400
774   K  A  V  R  I  G  Q  L  N  A  Y  Y  V  D  Y  K  N  V  T  N    793
2401 TTATAGTCTTAAAACCGCTGCTCAAGTAGGAGCGGACGATATCAAATAACTTACGCCACGG 2460
794   Y  S  L  K  T  A  A  Q  V  G  A  T  I  S  N  N  L  R  H  G    813
2461 ATTTGTTGACAATCAACAAGACGCATACACGCGCTTGGTTGCCAACTACTCTGATACGCG 2520
814   F  V  D  N  Q  D  A  Y  T  R  L  V  A  N  Y  S  D  T  R      833
2521 GAAGTGGATACGTGACAATTTTACATATAATTATAATATGGAGAAAGAAAAGTATAGGAT 2580
834   K  W  I  R  D  N  F  T  Y  N  Y  N  M  E  K  E  K  Y  R  I    853
2581 AAACCCAATACCACCATACACATGTGAGGTTGAAAGATTTGTTTCCATCCAGGAAAATAGT 2640
854   T  Q  Y  H  H  T  H  V  R  L  K  D  L  F  P  S  R  K  I  V    873
2641 TAAACTAGAGGGATATGAAGCCTTGTTGGCAATGATGCTAGACAGGTTTAACAACATAGA 2700
874   K  L  E  G  Y  E  A  L  L  A  M  M  L  D  R  F  N  N  I  E    893
2701 GTCAACACATGTAACTTTCTTCACATATTTAAGAGCACTACCTGACCGTGAAAAAGAAGT 2760
894   S  T  H  V  T  F  F  T  Y  L  R  A  L  P  D  R  E  K  E  V    913
2761 CTTTATTAGCTTAGTCTTAAACTATAATGGCCTTGGCAGAGAGTTGGTTGAAGTCTGAAGG 2820
914   F  I  S  L  V  L  N  Y  N  G  L  G  R  E  W  L  K  S  E  G    933
2821 TGTTAGGGCTAAACAAGCACAAGGTACTGTGAAATACGATATGAGTAAACTATTTGAACT 2880
934   V  R  A  K  Q  A  Q  G  T  V  K  Y  D  M  S  K  L  F  E  L    953
2881 GAATGTACTAGAGAACGGAGTTGACGAAGAAGTTGACTGGGAGAAAAGAGAAACGCAACAG 2940
954   N  V  L  E  N  G  V  D  E  D  V  W  E  K  E  K  R  N  R       973
2941 GTCAGATATCAAGACTGTTAACATAAGTTATGCAAAAGTTCTCGAACATTGTAGAGAGCT 3000
974   S  D  I  K  T  V  N  I  S  Y  A  K  V  L  E  H  C  R  E  L    993
3001 ATTCATCATGGCGAGGGCCGAAGGGAAACGGCCAATGAGGATGAAATGGCAAGAGTACTG 3060
994   F  I  M  A  R  A  E  G  K  R  P  M  R  M  K  W  Q  E  Y  W    1013
3061 GAGGCAGAGAGCAGTTATCATGCCAGGTGGATCGGTCCACAGTCAACATCCAGTCGAACA 3120
1014  R  Q  R  A  V  I  M  P  G  G  S  V  H  S  Q  H  P  V  E  Q    1033
3121 GGACGTGATTAGAGTATTACCCAGAGAAATCAGAAGTAAGAAGGGGGTGGCAAGTGTCAT 3180
1034  D  V  I  R  V  L  P  R  E  I  R  S  K  K  G  V  A  S  V  M    1053
3181 GCCATACAAAGAACAGAAGTATTTCACGTCCAGAAGGCCGGAAATACACGCTTACACTTC 3240
1054  P  Y  K  E  Q  K  Y  F  T  S  R  R  P  E  I  H  A  Y  T  S    1073
3241 AACGAAATACGAGTGGGGAAAAGTGAGGGCACTATATGGGTGTGATTTTTCATCACATAC 3300
1074  T  K  Y  E  W  G  K  V  R  A  L  Y  G  C  D  F  S  S  H  T    1093
```
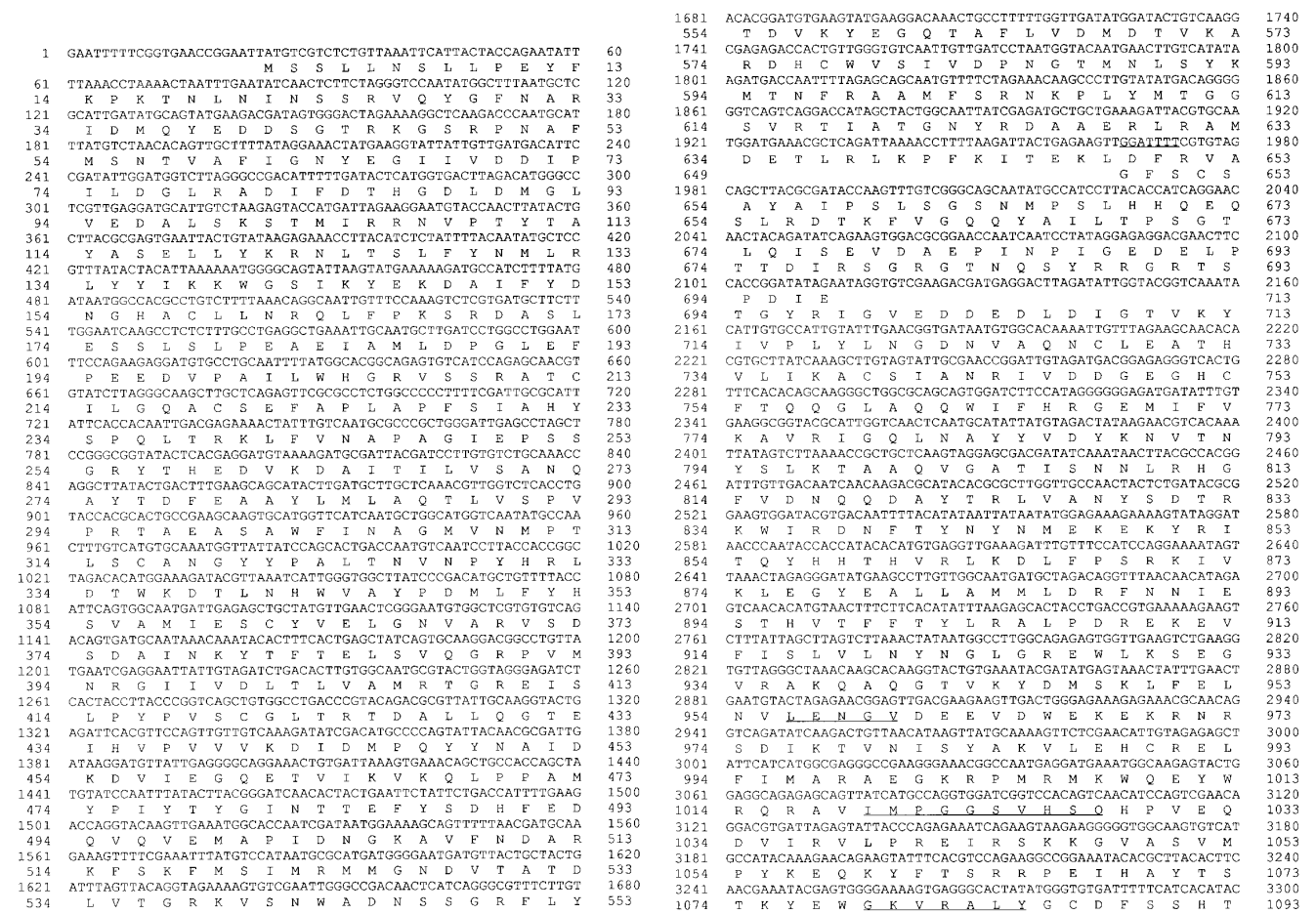
FIG. 2. Sequence of La. The complete sequence of the cDNA plus strand is shown, along with the predicted protein sequences. In the DNA sequence, the slippery site is underlined. In the protein sequence, the La sequences that correspond to the eight conserved motifs of the totivirus RDRPs are underlined. This cDNA sequence is deposited in Genbank under Accession No. U01060.

L1 (Fig. 1). There are two large overlapping reading frames, *cap* and *pol,* which in this case overlap for 153 bases. The *cap* ORF (bases 24–2114) predicts a protein (Cap) of 78.3 kDa and 697 amino acids in length. This ORF begins with the first AUG in the viral plus strand. The predicted size of La Cap (78.3 kDa) is quite close to that determined by SDS–PAGE, by which it has been estimated at 77 (*11*) and 80 kDa (*3*). The major *in vitro* translation product of denatured La dsRNA is a protein that comigrates with La Cap, as expected (*3*).

A *cap–pol* frameshift fusion protein of 1512 amino acids or 171.5 kDa is predicted, with a frameshift at a GGAUUUU slippery sequence starting at base 1967. Again, this is similar to the L1 Cap–Pol fusion protein of 171 kDa, which is generated by a frameshift at the slippery sequence starting at base 1957 (*5, 6, 12*).

The L1 packaging signal, two stems separated by an unpaired A residue, with several conserved bases in the loop, was not found in the La plus strand. This is consistent with the inability of ScVL1 particles to package La plus strands (*13*).

As described previously (*1*), the La and L1 RDRPs are 38% identical through the portion containing the eight conserved motifs in Pol. Overall, the Pol domains are 29% identical in predicted amino acid sequence, while the Cap domains are only 21% identical overall (below statistical significance).

Unexpectedly, there is an additional region of high conservation of sequence within the *cap* ORF (Fig. 3). This region of 60 amino acids is 37.7% identical in amino acid sequence in the two proteins. We previously noted a region of L1 Cap with some similarity to the picornavirus vp3, the most highly conserved of the picornavirus capsid polypeptides (*14*). The 60-amino-acid region of similarity between the L1 and La Cap proteins maps in the middle of the previously observed region of similarity between the L1 Cap and the picornavirus vp3. This lends further support to speculation that this region is involved in the protein–protein contacts necessary to assemble an icosahedron. Viral interference experiments are also consistent with this hypothesis (*15*).

In contrast, comparison of the RNA sequences of L1 and La detects only very limited similarities, confined to the most conserved region of *pol* (results not shown).

```
3301  AATGGCTGATTTTGGATTGTTACAATGCGAGGATACATTCCCGGGCTTTGTACCAACAGG  3360
1094   M  A  D  F  G  L  L  Q  C  E  D  T  F  P  G  F  V  P  T  G  1113
3361  GTCTTACGCCAATGAGGATTATGTCAGGACCAGAATTGCTGGGACTCACTCATTGATCCC  3420
1114   S  Y  A  N  E  D  Y  V  R  T  R  I  A  G  T  H  S  L  I  P  1133
3421  TTTCTGTTACGATTTCAACAGCCAACATTCAAAGGAAGCCATGCAAGCAGT  3480
1134   F  C  Y  D  F  D  D  F  N  S  Q  H  S  K  E  A  M  Q  A  V  1153
3481  GATTGATGCATGGATATCTGTCTATCACGATAAGTTAACAGATGACCAGATAGAGGCGGC  3540
1154   I  D  A  W  I  S  V  Y  H  D  K  L  T  D  D  Q  I  E  A  A  1173
3541  AAAGTGGACACGAAACTCGGTAGATAGAATGGTCGCTCACCAACCTAACACTGGTGAGAC  3600
1174   K  W  T  R  N  S  V  D  R  M  V  A  H  Q  P  N  T  G  E  T  1193
3601  TTATGATGTTAAAGGGACACTGTTTAGTGGCTGGCGATTAACAACATTTTTCAATACGGC  3660
1194   Y  D  V  K  G  T  L  F  S  G  W  R  L  T  T  F  F  N  T  A  1213
3661  GTTGAACTATTGCTACCTGGCTAATGCAGGTATAAACTCACTAGTGCCAACGAGTCTCCA  3720
1214   L  N  Y  C  Y  L  A  N  A  G  I  N  S  L  V  P  T  S  L  H  1233
3721  TAATGGTGATGATGTTTTTGCAGGGATAAGGACAATAGCTGACGGTATTTCTTTGATCAA  3780
1234   N  G  D  D  V  F  A  G  I  R  T  I  A  D  G  I  S  L  I  K  1253
3781  AAACGCCGCAGCCACGGGAGTTCGCGCTAATACAACTAAAATGAACATTGGTACGATAGC  3840
1254   N  A  A  A  T  G  V  R  A  N  T  T  K  M  N  I  G  T  I  A  1273
3841  AGAGTTTTTGAGAGTTGATATGCGTGCAAAAAATAGTACTGGCAGTCAGTATTTAACAAG  3900
1274   E  F  L  R  V  D  M  R  A  K  N  S  T  G  S  Q  Y  L  T  R  1293
3901  AGGGATTGCTACCTTCACGCACAGTAGGGTGAGTCTGATGCACCACTGACATTGCGCAA  3960
1294   G  I  A  T  F  T  H  S  R  V  E  S  D  A  P  L  T  L  R  N  1313
3961  TCTAGTATCTGCTTACAAAACCAGATATGACGAGATTTTAGCTCGTGGCGCAAGCATCGA  4020
1314   L  V  S  A  Y  K  T  R  Y  D  E  I  L  A  R  G  A  S  I  D  1333
4021  TAACATGAAGCCACTCTATCGTAAGCAATTATTTTTTGCTAGAAAGTTGTTCAATGTCGA  4080
1334   N  M  K  P  L  Y  R  K  Q  L  F  A  R  K  L  F  N  V  E  1353
4081  GAAGGACATTGTTGACAATCTGATAACGATGGACATATCATGTGGCGGTTTGCAAGAAAA  4140
1354   K  D  I  V  D  N  L  I  T  M  D  I  S  C  G  G  L  Q  E  K  1373
4141  GGGTAGGGTATCAGAGATGGTGTTACAGGAGGTTGACATTGACATAGATAGTTATAG  4200
1374   G  R  V  S  E  M  V  L  Q  E  V  D  I  E  N  I  D  S  Y  R  1393
4201  GAAGACAAGGATGATCGCCCAAACTGATTGACAAGGGGGTTGGCGATTATACTGCATTCCT  4260
1394   K  T  R  M  I  A  K  L  I  D  K  G  V  G  D  Y  T  A  F  L  1413
4261  GAAAACTAACTTTTCCGAGATAGCTGATGCTATCACAAGAGAGACACGTGTAGAGTCAGT  4320
1414   K  T  N  F  S  E  I  A  D  A  I  T  R  E  T  R  V  E  S  V  1433
4321  GACCAAGGCTTATAATGTTAAGAAGAAAACGGTCGTACGCGCGTTTAGGGACCTAAGCGC  4380
1434   T  K  A  Y  N  V  K  K  T  V  V  R  A  F  R  D  L  S  A  1453
4381  AGCATATCATGAAAGAGCGGTGAGACATGCTTGGAAGGGGATGAGTGGACTACACATAGT  4440
1454   A  Y  H  E  R  A  V  R  H  A  W  K  G  M  S  G  L  H  I  V  1473
4441  CAACAGGATTCGTATGGGAGTGAGCAACTTAGTAATGGTTGTTAGCAAAATCAATCCTGC  4500
1474   N  R  I  R  M  G  V  S  N  L  V  M  V  V  S  K  I  N  P  A  1493
4501  AAAAGCTAATGTGCTAGCCAAATCAGGAGATCCTACAAAATGGCTTGCAGTCCTTACATG  4560
1494   K  A  N  V  L  A  K  S  G  D  P  T  K  W  L  A  V  L  T  1512
4561  ATATACAGGCAACCACATAAGACCTGAGAACAAAGAGTACATACGATACTACGCA  4615
```

FIG. 2—Continued

Since ScVLa and ScVL1 coexist in the same cells, this very limited similarity, confined to a region known to be conserved throughout the totiviruses, implies that there is no recombination between the two viral genomes.

The similarity in genomic organization in L1 and La led us to expect a translational frameshift to generate an La Cap–Pol fusion product, as with L1. A −1 translational frameshift occurs on the L1 mRNA at the slippery site GGGUUUA, so that a Gly–Leu pair of tRNAs slips back a base onto the Gly–Phe codons, resulting in a protein with the sequence GLRS through the region of the frameshift (T.-H. Tzeng and J.A. Bruenn, unpublished), just as in the retroviruses (16). This event requires the presence of a downstream pseudoknot (17, 18), which causes a ribosomal pause at this site (19).

The putative slippery site of La is GGAUUUU. In this case, an Asp–Phe pair of tRNAs would slip back a base onto the Gly–Phe codons, resulting in a protein with the sequence DFSC through the frameshift site. This slippery site is similar to that of the MMTV gag–pro site, which is GGAUUUA, but this is the only known slippery site

```
275 ALRKYVNRNRLYNQFYTAAQLLAQIMMKPVPNCAEGYAWLMHDALVNIPK 324
    |:  .|. |. |.:| .| :|||.::.|||..|| ||::::.::||:|.
264 AITILVSANQAYTDFEAAYLMLAQTLVSPVPRTAEASAWFINAGMVNMPT 313

325 FGSIRGRYPFL 335
    ::: .| || |
314 LSCANGYYPAL 324
```

FIG. 3. The 60-amino-acid region of Cap in La and L1 that is similar (37.7% identity). Vertical lines indicate identity, double dots, similarity. This alignment was generated by GAP (24, 25). The La Cap sequence is the lower and the L1 Cap sequence the upper line.

TABLE 1

Frameshift Activity of La Minimal Region in Yeast

Region present in construct	Mutation	Activity	% read-through vector
Read-through vector	None	11,343	100
1957–2001	None	104.8	0.92
1957–2004	None	210	1.85
1957–2004	GGATTTT to GcATTTT	34.4	0.30
1957–2004	GGATTTT to GGATgTT	33.4	0.29

without a triplet of identical bases at the 5′ end (20). We have placed this region (the slippery site and adjacent sequences) in a β-gal reporter vector (pG4LacZ), which was derived from pG4 (15), by inserting LacZ as a NcoI flush end, KpnI sticky end fragment from p3p (21) into a SalI flush-ended–KpnI sticky-ended pG4. This allows us to detect frameshifting in yeast strain T120 (15) by the synthesis of β-gal in vivo (21). β-Galactosidase assays were performed as described (17) on whole cell lysates and expressed as arbitrary units per cell (22). All assays were performed on three independent transformants simultaneously. Standard deviations of measurements were usually less than 10% of the mean.

A minimal region of 48 bp does function, with a frameshift efficiency of about 2% with respect to the read-through vector (Table 1). We analyzed mutants (23) in which alterations of the slippery site prevent frameshifting (Table 1). The sixfold decrease in frameshift efficiency observed in these two slippery site mutants, in which a single base is altered in the first or second codon, is similar to that observed for two-base changes in the L1 slippery site (18). In short, the La pol ORF seems to be read by a translational frameshift, just as with L1.

ACKNOWLEDGMENTS

REFERENCES

1. Bruenn, J. A., Nucleic Acids Res. 21, 5667–5669 (1993).
2. Field, L. J., Bobek, L., Brennan, V., Reilly, J. D., and Bruenn, J., Cell 31, 193–200 (1982).
3. Reilly, J. D., Bruenn, J., and Held, W., Biochem. Biophys. Res. Commun. 121, 619–625 (1984).
4. Bruenn, J. A., In "Replication of dsRNA Mycoviruses" (R. D. J. Holland and P. Ahlquist, Eds.), pp. 195–209. CRC Press, Boca Raton, FL, 1988.
5. Diamond, M. E., Dowhanick, J. J., Nemeroff, M. E., Pietras, D. F., Tu, C.-L., and Bruenn, J. A., J. Virol. 63, 3983–3990 (1989).
6. Icho, T., and Wickner, R. B., J. Biol. Chem. 264, 6716–6723 (1989).

7. Franklin, R. M., *Proc. Natl. Acad. Sci. USA* **55**, 1504–1511 (1966).

8. Gubler, U., and Hoffman, B. J., *Gene* **25**, 263–269 (1983).

9. Sanger, F., Nicklen, S., and Coulson, A. R., *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467 (1977).

10. Bruenn, J. A., and Brennan, V. E., *Cell* **19**, 923–933 (1980).

11. Sommer, S. S., and Wickner, R. B., *Cell* **31**, 429–441 (1982).

12. Dinman, J. D., and Wickner, R. B., *Genetics* **136**, 75–86 (1994).

13. Yao, W., Muqtadir, K., and Bruenn, J. A., *J. Virol.* **69**, 1917–1919 (1995).

14. Bruenn, J. A., Diamond, M. E., and Dowhanick, J. J., *Nucleic Acids Res.* **17**, 7487–7493 (1989).

15. Yao, W., and Bruenn, J. A., *Virology* **214**, 215–221 (1995).

16. Jacks, T., and Varmus, H. E., *Science* **230**, 1237–1242 (1985).

17. Tzeng, T.-H., Tu, C., and Bruenn, J. A., *J. Virol.* **66**, 999–1006 (1992).

18. Dinman, J. D., Icho, T., and Wickner, R. B., *Proc. Natl. Acad. Sci. USA* **88**, 174–178 (1991).

19. Tu, C.-L., Tzeng, T.-H., and Bruenn, J. A., *Proc. Natl. Acad. Sci. USA* **89**, 8636–8640 (1992).

20. ten Dam, E. B., Pleij, C. W. A., and Bosch, L., *Virus Genes* **4**, 121–136 (1990).

21. Belcourt, M. F., and Farabaugh, P. J., *Cell* **62**, 339–352 (1990).

22. Miller, J. H., "Experiments in Molecular Genetics." Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1972.

23. Kunkel, T. A., *Proc. Natl. Acad. Sci. USA* **82**, 488–492 (1985).

24. Devereux, J., Haeberli, P., and Smithies, O., *Nucleic Acids Res.* **12**, 387–395 (1984).

25. Genetics Computer Group, "Program Manual for the GCG Package, Version 7." GCG, Madison, WI, 1991.