Structure and expression of a second sea urchin U1 RNA gene repeat

Jin-Chen Yu, Micheal A.Nash, Carlos Santiago, William F.Marzluff

Department of Chemistry, Florida State University, Tallahassee, FL 32306, USA

Received 12 September 1986; Revised and Accepted 14 November 1986

ABSTRACT

There are two tandemly repeated sets of U1 RNA genes in the sea urchin L. variegatus. Each of these genes is present in a 1.4 kb repeat defined by a HindIII site about 450 bases 5' to the gene. The sequences of a member of both repeating units (U1.1 and U1.2) has been determined. The repeats are nearly identical for 550 nucleotides 5' to the gene but show great divergence starting 30 nucleotides 3' to the gene, just after the CAAAGAAAGAAAA sequence thought to be required for 3' end formation. The other boundary between the conserved and non-conserved sequences is a polypyrimidine sequence (on the strand which codes for U1 RNA). Both of these repeats are expressed in blastula stage embryos, as judged by transcription of unique sequences 3' to the gene in nuclei isolated from blastula stage embryos. At least some of the two types of repeats are interspersed, since representatives of both repeat types on a single phage isolated from a gene library. The sequence of the U1 RNA in L. variegatus eggs and embryos corresponds to the sequence of the U1 repeat.

INTRODUCTION

Multiple copies of the U1 RNA gene are present in all organisms in which these genes have been studied (1-6). In birds and mammals the genes are not present in an obvious repeating unit (1-3), although the possibility of a large repeated unit cannot be ruled out for humans since the genes isolated thus far are virtually identical in both the 5' and 3' flanking regions (4). The human U2 genes are organized in a large repeated unit (7). In chickens the U1 genes are not tandemly repeated (8) and the rat U1a genes and mouse U1b genes are also not present in a tandemly repeated unit (5,6). In Xenopus some of the U1 genes are present in tandemly repeated units (9, 10). The Xenopus repeat units identified thus far each contain two U1 genes which code for U1 RNAs with different sequences (10). There are also other U1 genes in Xenopus which are not present in the major tandemly repeated units (11).

The expression of the U1 genes in Xenopus is regulated during development (9, 11). There is developmentally regulated expression of the mouse U1

RNAs as well (12). The U1b RNAs are expressed in early mouse development and in many tumor lines but not in many terminally differentiated cells (12) or in fibroblasts (13). Whether or not the different U1 RNA sequence variants are functionally different is not known.

During development of the sea urchin embryo there are many changes in the expression of genes coding for nuclear components, including the histones (14) and the nuclear lamins (15). There are dramatic changes in the rates of expression of the small nuclear RNA genes as well (16). We have previously reported that most of the genes coding for U1 RNA are present in a 1.4kb tandemly repeated unit in the sea urchin *L. variegatus* (17). One copy of the U1 gene is present in each repeating unit. The repeat unit is defined by single Hind III and HincII sites flanking the gene, and a BglII site inside the gene. We also reported that there were at least two different members of the repeat defined by differences in restriction sites in the region 3' to the U1 gene (18). We report here the isolation and structure of a member of the second 1.4 kb repeat class. As predicted from the restriction maps the major differences between the two repeats lie in the region 3' to the U1 gene. One phage isolated from a gene library contains both types of repeat units indicating that at least some of the two types of repeat units are interspersed. Utilizing the fact that the regions 3' to the U1 gene are transcribed in isolated nuclei from sea urchin embryos (19), we show that both of these repeats are expressed in blastula embryos at rates proportional to the gene number. The sequence of the U1 RNA found in the sea urchin egg is identical with the sequence predicted from these repeat units, consistent with the possibility that these genes are also active during oogenesis.

MATERIALS AND METHODS

Isolation and sequence of pLvU1.2. Sperm DNA from *L. variegatus* was digested with HindIII and DNA fragments 1.2-1.6 kb in length were isolated by agarose gel electrophoresis. These were cloned into the HindIII site of pUC9. The white ampicillin resistant colonies were screened using the nick-translated 1.4 kb insert of pLvU1.1 which contains a *L. variegatus* U1.1 gene. Two positive colonies were obtained, which had identical inserts as judged by restriction enzyme mapping with 10 enzymes. The sequence of one of these inserts was determined by the method of Maxam and Gilbert (20). The sequence of both strands was obtained. The sequence through all the restriction enzyme sites used in sequencing was obtained directly.

Subcloning of unique regions of pLvU1.1 and pLvU1.2

The BglIII-HindIII fragment from pLvU1.1 and pLvU1.2 containing the region 3' to the gene was subcloned into pUC18. The HphI to TaqI fragment from pLvU1.1 (nucleotides 68 to 240 3' of the end of U1 RNA) was cloned into the AccI-SmaI site of pUC18. The AvaII to RsaI fragment from pLvU1.2 (nucleotides 100 to 250 3' of the end of U1 RNA) was cloned into the SmaI site of pUC18. These subclones were excised from the plasmid with EcoRI and HindIII and introduced into the M13 vectors, mp8 and mp9. These M13 phage yielded single-stranded probes for both strands of the DNA.

Sequence of U1 RNA

The sequence of U1 RNA in sea urchin eggs and embryos was determined by extending a ^{32}P labeled oligonucleotide primer with reverse transcriptase in the presence of dideoxynucleotides. The conditions used were those of Nash et al. (21) except that the concentration of deoxynucleotides was increased to 100 M, keeping the ratio of dideoxynucleotides to deoxynucleotides constant. The primer was synthesized using an Applied Biosystems DNA Synthesizer and was a generous gift of Dr. Don Sittman.

Transcription of the U1 RNA gene in isolated nuclei

Nuclei were prepared from blastula embryos just after hatching as described by Morris and Marzluff (22). Nuclei were incubated in the presence of α - ^{32}P CTP for 30 minutes at 25° and RNA was prepared as previously described (22). The RNA was hybridized with the single-stranded M13 DNA probes from both pLvU1.1 and pLvU1.2 immobilized on nitrocellulose. The filters were washed, exposed to X-ray film and the hybridized RNA detected by radioautography.

RESULTS

Previously we reported the isolation from L. variegatus sperm DNA of a 1.4kb fragment derived from a tandemly repeated unit containing U1 RNA genes. The sequence of this fragment has been reported (18). Using this fragment to probe the U1 RNA gene organization in L. variegatus we discovered that there were at least two types of 1.4 kb HindIII repeat units, which differed in several restriction enzyme sites, and that the gene we had isolated was characteristic of the minor repeat, about 20% of the total (18). We therefore have screened 1200 additional colonies containing plasmids with 1.2-1.6 kb HindIII fragments and have isolated a second plasmid, pLvU1.2, containing a U1 gene. The restriction map of this plasmid was consistent with it being a member of the major L. variegatus U1 repeat.

A

TWO TYPES OF U1 REPEAT

B

U1-1 -644 AATCTAT AAAAGGTAAAA TATCCAGGTGATGTCATTTCACAGGGGTTG (+642)	53 TGGTGCACATGAGTGGACATATCCCTCCCAA TTTTGGGAAATTCACCTTATCAGG I
U1-2 -619 GATCTTTTAAAAAGAAAAAATATCAGGGTGCATGGTTATTTCCTAATGGGTTG (+632)	61 TGGGATATATCCCTTCAAACATTTGCTCAATTCAGCTTATCATGCT
-595 GAGTCCAGGTATTGACACTCATTTAATGTATGAACATAGTCTTTTGTGTGAT (+702)	111 TTTTCAACACACATAGTTCCTTATTTA CTTAAGGGGACATGATTGGC
-564 TCGTCTAGGCTACGACTATAA CCGTCTGAACTAGTTTCACTTTATGTTG CAT (+685)	113 TTTTATGACATATGGTGAATTTTCAACCAATTTGTTCTTTAGGGGGTTAGATTTTA
-535 TTT CTTTAT CCTTGATATGGAATAAACAATCTTTTAACTTATCAGATATT (+759)	157 ATTCCTTC ANTAAGG CTCGGGTTAAATCCGTGTGACAAAAAT ABCATACACTT
-513 TTTTCTTTTTCCTGCTGATAT GAATAAACAATCTTTTAACTTCATAGAGATATC (+742)	173 ATTCCTTCATGAAAAATCA GACTAAATGATGACAT
-478 AAGCTTCACA TTTCACTTAAATGAGACAAAGTCTTCGGGTCTACTCAACATAGA	214 TACACTCTGCTCATCTACCT GCGAATGTATGT CGAATTAACCT
-453 AAGGCTCACACATTTTCACTTAAATGGGACGAAGCTT ACCCACTAGA	218 CCGTAAATGTTGGCCATGTACACAGCGAAATGTGTTCCAAATAATATTTGTAATTT
-420 GCCCATGGAACGAGATATTAAAGAGACAAAGTATTAATTTTCCCAAGTAAGACGACGAC	261 TGCATGTTTGAAGATCTAATGAGATTTAAAAAAATGATTTCGGGAGATGGCT
-404 GCCAATGGAACGAGATATTAAAGAGAGGATGAT T GACGACGAC	278 AGAAAAACAATCCGACTCTTAGTGGCT
-360 CGATAAAGAAAGGTTGACCCGATAACAGGATGAGTAAGCTGACGCTTATGTCTATT	321 GATCTATTACTGCTGCGAAGTTTCCGATAAAACCCCTCCCAATAGTTTCTTCATT
CAATAAAGAAAGGTTGACCCGATAACAGGATGAGTAAGCTGACGCTTATGTCTATT	308 TGCCTACTTGTCTGGTGCAGAAATTCGATAAAG GAGATTTAAACAGTTTCTTATT
-300 AATCTGCAAGCTCAGTAAAGTTATGTGAGTCTCATGATATCAACCTCTACTCTTC	381 GAACAACGCTCATCAAAATACGTTTCAAGTAACTCTCAATCAATACATATTTGTAAC
AACTCTTAACTGCTCAGTAAAGTTATGTGAGTCTCATGATATCAACCTCTACTCTTC	367 GACACATATCTGTAATAATACCTTTGATGGTAAATTTTCTCAAGATGCTTATTCAC
-240 ACATCTTACGTGCAATTTTGTACTGTATTATGTTTAAATATAGATATACATGGATAC	441 TGGGTATTTTAAACCTATTTGGCCATCTGTTGTT CTTTCCATATGCAATAGTT
ACATCTTACGTGCAATTTTGTACTGTATTATGTTTAAATATAGATATACATGGATAC	427 TTGGTATCTCTCAAACTTATGGTTTACGTTTGTATGTTTACGTTTATGATGACATGGTT
-180 GTCAATTAATGTTGCTCCAGAGAGGGGACAGAGATTTGAGGTTCTTCTCTCAAACTT	509 AGTGGTAAAAAGTTGATGATGATCTGATCACTAAAAAATAAATGACCTTTAAGG
GTCAATTAATGTTGCTCCAGAGAGGGGACAGAGATTTGAGGTTCTTCTCTCAAACTT	487 ACCGTGATGATGTTGATATGATGCTCACTAAAAA CAAAGATCTGTTTAA G
-120 GCCAAATTTGACAATTCATAAATCATGTGACGTTTCTCTCTTGTAGTGTGAGATGT	560 GTCAAGCTTGGACCAACCAATCTTCTGTCACCAATCTAT AAAAGGTAAAA TAT
GCCAAATTTGACAATTCATAAATCATGTGACGTTTCTCTCTTGTAGTGTGAGATGT	544 GTCAAGCTTTACCAACCACTGCTTCTGTCACCACTTTTAAAAAGAAAAAATAAT
-60 TTAATCTCTTCTCTGAGATAGATTTGATCAAGTACGCGGAGGATGATGAGGAGC	614 CCAGGTGCTATGCTTCTCCAGGGTGGAGTCCAGGTTATGACTACTATTATGTTG (-565)
TTAATCTCTTCTCTGAGATAGATTTGATCAAGTACGCGGAGGATGATGAGGAGC	604 CAGGGTGCATGGTTATTTCCCAATGGGTGCTGCTAGGCTACGACTATAA CG (-541)
1 ATATCTTACCTGCGCAGGGTGCATTTGATCAAGAGATGACACCCCAAGGCGAGGCTT	674 TATGACTAGTTTGGATTTTGGTGGCTTTTCTTTAT CCTTGTATATGGAATA (-508)
ATATCTTACCTGCGCAGGGTGCATTTGATCAAGAGATGACACCCCAAGGCGAGGCTT	657 TCTAAGTATGTTTCAATTTGTGG CATTCTTTTCTCTGATAT GAAATA (-484)
61 GCAATCTGACCTGCGCTTCTGAACTTCTGATTTCCCAACCTGCGGAGCTGCGGCT	731 AACAATCTTTTAACTTATCAGATATTAAGCTCACA TTTCTTCTTAAATGAGAC (-505)
GCAATCTGACCTGCGCTTCTGAACTTCTGATTTCCCAACCTGCGGAGCTGCGGCT	714 AACAATCTTTTAACTTATCAGATATCAAGCTCAGCTTCTTCTTAAATGAGAC (-424)
121 ACAAATTATGATAGCGAGATCTGCTTCTGCTATCTCT 161	
ACAAATTATGATAGCGAGATCTGCTTCTGCTATCTCT	

Fig. 1. The U1 repeats of *L. variegatus*. A. The restriction map of the repeat units cloned in pLvu1.1 and pLvu1.2 are shown as well as the sequencing strategy used to sequence the U1.2 repeat. The * indicate 5' end-labeled fragments and the 3' end-labeled fragments. B. The sequences of pLvu1.1 and pLvu1.2 are compared. Base 1 is the first base of the U1 RNA. The HindIII site is from -446 to -441 (U1.1) and from -419 to -414 (U1.2). The RNA sequence is in bold face and the polypyrimidine tract at the borders of the homologous regions is underlined. The sequences are aligned to give maximum homology with the minimum number of insertions and deletions. This comparison has substantial overlap at both the beginning and the end to better show the breaks in homology regions. The negative numbers indicate the distance 5' to the U1 RNA and the positive numbers distance 3' of the U1 RNA. The places where the sequences differ are marked with an *.

Comparison of the two repeat units

Figure 1A shows the restriction map and Figure 1B the sequence of the pLvU1.2 insert compared with the previously reported pLvU1.1. The pLvU1.2 insert contains 1356 nucleotides, 42 nucleotides shorter than pLvU1.1. There is a striking non-random pattern of divergence between these two repeat units. The region 5' of the U1 gene, including the HindIII site and 100 bases further 5' is highly conserved (>95%) between the two genes. The major differences in this region are two deletions in the pLvU1.2 clone which largely account for the difference in length between the two clones. This region is bordered by a pyrimidine rich sequence in both clones.

The homology between the two repeats ends just 3' of the CAAAGAAAGAAAA sequence 3' of the gene which may be part of the signal for proper 3' end formation (19). Up to this point the two clones are identical, including the 15 bases between the end of the U1 RNA and the CAAAGAAAGAAAA sequence, except for a single change at nucleotide 131 in the U1 RNA. The two repeats are highly divergent from the CAAAGAAAGAAAA sequence to the pyrimidine rich region which marks the end of the 5' region of homology. There are small regions within the 3' sequence which retain some homology, but the repeats have clearly diverged greatly with both nucleotide changes and insertions and deletions having occurred in the 3' spacer region. This is the region where sequence difference were predicted from the Southern blots of genomic DNA using different restriction enzymes (18).

Some of the U1.1 and U1.2 units are interspersed

Using the U1 coding region as a probe, we screened a λ phage gene library constructed by partial Sau3A digestion of sperm DNA. Among the 37 phage isolated was one which we have identified as an end fragment of a tandemly repeated cluster. Partial sequence of two of the 1.4 kb HindIII fragments revealed that one of these had the 3' flanking sequence characteristic of the U1.1 repeat unit and the other the 3' flanking sequence characteristic of U1.2 (Fig. 2). There are several base substitutions in each sequence which are presumably polymorphisms, but the basic sequence of each one is clearly recognizable. Thus the two types of U1 repeats are closely linked. It is not known whether this is a characteristic of the end fragments of the repeat or whether they are interspersed in a uniform pattern.

Both the genes are transcribed in sea urchin embryos

Since the two types of repeat units code for the same RNA, it is difficult to assess whether they are each expressed. RNA polymerase II transcribes into the 3' flanking regions of most genes and transcription of individual

Fig. 2. The U1.1 and U1.2 genes are linked. The partial sequence of two 1.4 kb HindIII fragments subcloned from the phage LvU1D is shown starting at nucleotide 121 of the coding region and extending 3' to the gene. The sequences are aligned as in Fig. 2 and the *'s indicate the differences noted in Fig. 1B. U1.1A had the same sequence as U1.1 and U1.2A the same sequence as U1.2 except where shown.

members of a gene family can be measured by measuring transcription of the flanking region in isolated nuclei (23). We have previously shown that nuclei isolated from sea urchin embryos actively transcribe the U1 RNA (22) genes. Using a DNA-dependent system we have shown that transcription extends 3' of the gene in vitro (19). The evidence presented below suggests that in isolated sea urchin nuclei transcription proceeds 3' to the U1 gene at least some of the time.

Since the regions 3' of the two U1 RNA genes differ dramatically we reasoned that transcription of specific genes might be measured by measuring transcription of these regions. The subclones shown in Fig. 3A were constructed from pLvU1.1 and pLvU1.2. The probes are derived from regions just 3' of the gene which are transcribed in vitro in a DNA-dependent system (19).

When RNA synthesized in isolated nuclei was hybridized to the single-stranded probes, the results shown in Fig. 3B were obtained. There was hybridization to the coding strand, but not the non-coding strand of each gene. There was more hybridization to the pLvU1.2 clone than to the pLvU1.1 clone, in agreement with the relative number of copies of each type of repeat unit in the *L. variegatus* genome (18).

It is not known whether these regions are normally transcribed. The relative amounts of transcription (U1.2 > U1.1) is likely accurate since the U1.1 probe starts closer to the actual gene, but this interpretation depends on the assumption that the extent of transcription of the 3' flanking region

Fig. 3. Expression of the U1.1 and U1.2 repeats. A. Fragments A and B specific for the U1.1 and U1.2 repeats respectively were subcloned into mp8 and mp9 to generate single-strand probes. Subclone A is from nucleotides 68 to 240 3' of the end of the U1 RNA in pLvU1.1 and subclone B is from nucleotide 100 to 240 3' of the end of the U1 RNA in pLvU1.2. Subclone C is the previously described coding region subclone from nucleotide 30 to 140 of U1 RNA, mpU1C (22). B. RNA synthesized in isolated nuclei in the presence of ^{32}P CTP was hybridized to the immobilized DNA of the subclones shown in A. The filters were washed and the hybridized RNA detected by autoradiography.

is the same for each type of repeat unit. Similar results, more transcription of the U1.2 flanking region than the U1.1 flanking region have been found in 12 different nuclear preparations. We conclude that both of these repeats are actively transcribed in the hatching blastula embryos, probably at about the same rate per gene.

Sequence of *L. variegatus* U1 RNA

Since the U1.1 and U1.2 repeat units we cloned differed by a single base in the U1 RNA, we determined the partial sequence of U1 RNA in the egg and gastrula embryo of *L. variegatus* (Fig. 4). The sequence of the U1 RNA in the

Fig. 4. Sequence of the *L. variegatus* U1 RNA. A. The sequence of the U1 RNA was determined by extending an oligonucleotide primer with dideoxynucleotides. Left: gastrula RNA; Right: egg RNA. The sequence of the RNA from nucleotide 115 to 133 (top to bottom) is shown next to the gel. The numbers refer to the position in the U1 RNA sequence. Some bases could not be read because the reverse transcriptase paused presumably due to RNA secondary structure. At other positions there were reproducibly two bases at the same position. B. The sequence of the U1 RNA is compared to the gene sequence. An * indicates that the base could not be read. The dashes indicate the primer used for sequencing. Two bases are indicated at positions of apparent heterogeneity. The RNA sequence is a composite of several independent sequences. Note that position 123 is a G/A in the gastrula sequence shown, but was an A in the egg sequence and could be unambiguously read as an A in other egg and gastrula sequences (not shown). Position 122 was reproducibly a mixture of C and U.

gastrula embryo should be characteristic of that of the RNA synthesized in morula and blastula, the time of the most intense U1 RNA synthesis (16). The sequence was obtained using a synthetic oligonucleotide complementary to the 3' end of the RNA. The oligonucleotide was end-labeled and then hybridized

with total egg RNA or with total nuclear RNA from gastrula embryos. The hybrids were isolated and extended with AMV reverse transcriptase in the presence of dideoxynucleotides. The four reactions were analyzed by gel electrophoresis and the RNA sequence read directly from the gel (Fig. 4A). The sequence of all but 3 nucleotides of the U1 RNA (except for the region complementary to the primer) could be determined in this way (Fig. 4B). Occasionally there were places where the reverse transcriptase paused, presumably due to secondary structure or base modifications, and the sequence at that nucleotide could not be determined. There are other places where reproducibly there were two lanes which contain bands at a particular position, which may represent heterogeneity in the RNA sequence. For example, at position 122 there is reproducibly (6 different sequences with 3 different RNA preparations) both a C and a U in the RNA sequence while both genes code for a C at this position (Fig. 4A). This suggests that other U1 genes in L. variegatus will have a T in this position. At other positions (e.g. 118, 123 and 128 in the gels shown for the gastrula RNA) there was apparent heterogeneity in some sequences but in the majority there was only a single nucleotide. The sequence of the egg U1 RNA is identical with that of gastrula U1 RNA in all the positions determined (Fig. 4B).

The sequence of the U1 RNA matches that of the U1.1 clone with a G at position 131 and no hint of an A at this position found in the U1.2 clone (Fig. 4A). This suggests that the U1.2 clone we obtained is a minor polymorphism in this repeat unit. This is supported by the sequence of the U1.2A clone shown in Fig. 2 which has an A at position 131.

DISCUSSION

The tandemly repeated U1 RNA genes in the sea urchin L. variegatus fall into two major classes as judged by Southern mapping with different restriction enzymes (18). These two types of repeat unit are about the same size and are characterized by a single HindIII site flanking the gene. The two U1 repeat units show an unusual pattern of sequence conservation; nearly exact conservation of the sequence 5' to the gene for 400 bases except for two small deletions in pLvU1.2 and very large divergence 3' to the gene starting immediately after the signal for 3' end formation. Because of the nature of tandemly repeated genes, these two regions meet at a polypyrimidine sequence, which marks the start of the conserved region. The two types of repeating units are intermingled with U1.1 and U1.2 type repeat units found within 12 kb of each other in a phage from a gene library. The exact organization of the

two types of repeats is not known. That the two types of repeats are found in the same relative proportion in six different individual sea urchins (our unpublished results) suggests that the overall organization of the U1 genes is stable.

There are two possible explanations for this unusual pattern: either there is strong functional conservation of the 5' flanking region for a long distance or there has been gene conversion maintaining the homology of this region of the repeat unit. If there has been gene conversion within a portion of the repeat unit, the polypyrimidine sequence may mark one end of the gene conversion unit. The other end may be at the CAAAGAAAGAAAA sequence at the end of the homology region or possibly elsewhere in the gene since the 3' border of homology corresponds with the presumed end of a functional sequence. The fact that there are no changes in the two units between the end of the gene and the CAAAGAAAGAAAA sequence, a region which may not be under as strong selective pressure as the gene region, suggests that the CAAAGAAAGAAAA sequence may indeed be a border of the gene conversion event. It is possible that there is intense selective pressure on the regions 5' to the sea urchin genes but we view this as less likely. In support of this possibility only 200 nucleotides 5' to the gene is required for maximal expression in vitro (19, and unpublished results). However, there are large regions of homology 5' to mammalian small nuclear RNA genes, some of which are conserved among different species (4-6).

Expression of the U1.1 and U1.2 genes

These genes code for an RNA which is identical in sequence to the U1 RNA in both the egg and gastrula nuclei. Both of these genes are expressed during the intense period of U1 synthesis in the sea urchin embryo, as judged by transcription of the divergent 3' flanking regions in isolated nuclei from sea urchin embryos. The finding that transcription extends 3' to the U1 gene in isolated nuclei may not necessarily reflect what happens in vivo. Recent evidence suggests that in the mammalian U1 genes 3' end formation occurs coincident with transcription, possibly as transcription termination (24, 25). Transcription into the flanking region in isolated sea urchin embryo nuclei may reflect inefficient 3' end formation, transcription which continues after 3' end formation or a difference between mammalian and sea urchin U1 RNA genes. Since the U1 coding regions are essentially the same, the observed transcription into the 3' flanking region allows us to conclude that some members of both of these types of repeat units are transcribed in blastula embryos.

Both these repeats may also both be expressed in oogenesis, another

period of intense U1 RNA synthesis which may require multiple gene copies. Thus it is likely that these two repeat units are functionally equivalent, like the different repeat units of the early histone genes found in L. pictus (26, 27). The two U1 repeats may have diverged from each other a long time ago, but have maintained a partial homology by gene conversion involving half of the repeat unit. The fact that at least some members of the two types of repeats are linked suggests that simple gene duplication and gene conversion cannot account for the structure of a stretch of tandemly repeated U1 genes. This is different from the early histone gene repeats which have undergone relatively uniform divergence in the flanking regions while maintaining high conservation in the coding regions presumably by selective pressure (26-28).

It is likely that there are other U1 genes present in L. variegatus other than those reported here. We have recently found that the sequence of the U1 RNA in somatic tissue differs in at least one base from that in eggs and embryos (M. Nash, unpublished results) indicating that another gene set must be expressed in somatic tissue. Currently we are isolating other members of the sea urchin U1 gene family which may be expressed at other developmental stages.

ACKNOWLEDGEMENTS

This work was supported by NIH grant GM 27789 to W.F.M. C.S. is supported by an NIH postdoctoral fellowship, GM 10301. We thank Dr. Don Sittman for synthesis of the oligonucleotide primer, Drs. J. Beidler and R. Rill for the L. variegatus library and Dr. Gil Morris for helpful discussions. This is publication #1036 from the Florida State University Marine Laboratory.

REFERENCES

1. Lund, E. and Dahlberg, J.E. (1984) J. Biol. Chem. 259 2013-2021.
2. Bernstein, L. B., Manser, T. and Weiner, A. (1985) Mol. Cell Biol. 5 2159-71.
3. Roop, D.R., Kristo, P., Stumph, W.E., Tsai, M.-J., and O'Malley, B.W. (1981). Cell 23 671-680.
4. Manser, T. and Gesteland, R.F. (1982). Cell 29 257-64.
5. Watanabe-Nagasu, N., Itoh, Y., Tani, T., Okano, K., Koga, N., Okada, N., Ohshima, Y. (1983) Nuc. Acids Res. 11 1791-1801.
6. Marzluff, W.F., Brown, D.T., Lobo, S.M., Wang, S.-S. (1983) Nuc. Acid Res. 11, 6255-6270.
7. Ares, M., Mangin, M. and Weiner, A.M. (1985). Mol. Cell. Biol. 5 1560-1570.
8. Earley, J.M., Roebuck, K.A. and Stumph W.E. (1985). Nuc. Acids Res. 12 7411-21.
9. Lund E., Dahlberg, J.E., Forbes, D.J. (1984) Mol. Cell Biol. 4 2580-86.
10. Zeller, R., Carri, M.T., Mattaj, I.W. and DeRobertis, E.M. (1984) EMBO J. 3 1075-1081.

11. Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984). *Cell* 38 681-689.
12. Lund, E., Kafan, B. and Dahlberg, J.E. (1985). *Science* 229, 1271-1274.
13. Moussa, N.M., Lobo, S. M. and Marzluff, W.F. (1985) *Gene* 36 311-319.
14. Newrock, K.M., Cohen, L.H., Hendricks, M.B., Donnelly, R.J. and Weinberg, E.S. (1978) *Cell* 14 327-36.
15. Schatten, G., Maull, G.G., Schatten, H., Chaly, N., Simerly, C., Balczon, R., and Brown, D.L. (1985) *Proc Nat. Acad. Sci.* 82 4727-31.
16. Nijhawan, P.N. and Marzluff, W.F. (1979). *Biochemistry* 18, 1353-1360.
17. Card, C.O., Morris, G.F., Brown, D.T. and Marzluff, W.F. (1983) *Nuc. Acids Res.* 10 7677-7688.
18. Brown, D.T., Morris, G.F., Chodchoy, N., Sprecher, C. and Marzluff, W.F. (1984) *Nuc. Acids Res.* 13, 537-56.
19. Morris, G.M., Price, D.H. and Marzluff, W.F. (1986) *Proc. Nat. Acad. Sci.* 83 3674-78.
20. Maxam, A. M. and Gilbert, W. (1980) *Methods in Enzymology* 65 499-560.
21. Nash, M.A., Brizzard, B.L., Wong, J.L. and Murphy, E.C. (1985) *J. Virol* 53 624-33.
22. Morris, G.F. and Marzluff, W.F. (1985) *Mol Cell. Biol.* 5 1143-50.
23. Chodchoy, N., Levine, B.J., Sprecher, C.S., Skoultchi, A.I., and Marzluff, W.F. (1987) *Mol. Cell. Biol.* 7, in press.
24. Hernandez, N. and Weiner, A. (1986) *Cell* 47 249-258.
25. Neuman de Vegvar, H. E., Lund, E. and Dahlberg, J. E. (1986) *Cell* 47 259-66.
26. Cohn, R.H. and Kedes, L.H. (1979) *Cell* 18 843-53.
27. Roberts, S.B., Weissner, K.E. and Childs, G. (1984) *J. Mol. Biol.* 174 647-62.
28. Holt, C.A. and Childs, G. (1984) *Nuc. Acids Res.* 12 6455-6471.