

Organization and Expression of Multiple Actin Genes in the Sea Urchin

RICHARD H. SCHELLER,¹† LINDA B. McALLISTER,¹ WILLIAM R. CRAIN, JR.,²
DAVID S. DURICA,² JAMES W. POSAKONY,¹ TERRY L. THOMAS,¹ ROY J. BRITTEN,¹
AND ERIC H. DAVIDSON¹*

Division of Biology, California Institute of Technology, Pasadena, California 91125,¹ and The Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545²

Received 3 March 1981/Accepted 29 April 1981

A set of at least 11 actin genes has been isolated from genomic recombinant deoxyribonucleic acid libraries of the sea urchin *Strongylocentrotus purpuratus*. Most of the isolates derive from a library which represents the genome of a single animal. There are at least five distinct types of sea urchin actin gene, some of which are represented by multiple copies in the genome. The actin gene types are distinguished by nonhomologous flanking sequences and intervening sequences, though the protein coding sequences appear in most cases to be quite similar. Eight of the 11 genes isolated have been recovered in λ recombinants that contain two actin genes, linked at 5- to 9-kilobase distances. Restriction map overlaps suggest that the genome contains an array of at least three of these genes spaced over about 30 kilobases of deoxyribonucleic acid. In the linkage patterns observed, actin genes of diverse types were linked to each other. In early embryos, actin messenger ribonucleic acid (RNA) transcripts of 1.8 and 2.2 kilobases were found, and the longer of these transcripts was more prevalent in the maternal RNA of the egg. From RNA gel blot experiments, we conclude that the two transcripts derive from different actin gene types. Different repetitive sequences were located to either side of most of the actin genes, and in most observed cases the repeat sequences which were adjacent to actin genes of a given type were similar. The repeat sequences flanking the actin genes belonged to families which were transcribed, but those repeats in the neighborhood of the actin genes which have been investigated were not themselves represented in the stable RNAs of eggs or early embryos.

Multiple actin genes have been recovered from genomic recombinant deoxyribonucleic acid (DNA) libraries of *Drosophila melanogaster* (13a, 14, 34), of the cellular slime mold *Dictyostelium discoideum* (19, 28), and of the sea urchin *Strongylocentrotus purpuratus* (11; M. A. Schuler and E. B. Keller, *Nucleic Acids Res.*, in press). Durica et al. concluded that there are between 5 and 20 actin genes in the genome of this sea urchin (11). In the present study we describe several actin complementary DNA (cDNA) clones and present evidence for at least 11 nonallelic actin genes, all of which have been isolated on cloned fragments of genomic *S. purpuratus* DNA. We believe this collection to include most, though perhaps not all, of the actin genes of this species. The majority of the actin genes we have characterized derive from a recombinant λ -genome library constructed from

the sperm DNA of a single individual. This turns out to have been of critical importance, in that in several cases it permitted us to distinguish between allelic polymorphism and gene multiplicity.

The picture of the organization of the sea urchin actin gene family which results from these studies is complex. Four main aspects are dealt with in this paper. These are the existence of several different actin gene types, distinguished by nonhomologous flanking and intervening sequences; the frequent occurrence of physical linkage between actin genes of different types; the presence of specific sets of repetitive sequences which flank the actin genes; and the expression of at least two different types of actin genes during early sea urchin development. As described elsewhere by Crain et al. (W. R. Crain, Jr., D. S. Durica, and K. Van Doren, *Mol. Cell Biol.*, in press), and also demonstrated here, at least two species of actin gene transcript can be

† Present address: Institute of Cancer Research, College of Physicians and Surgeons, New York, NY 10032.

distinguished in sea urchin embryos, and these species differ in the temporal pattern of their expression. Because the number of genomic sea urchin actin genes is so large and their nature is diverse, the following account is by no means complete. The overall features of the organization of the actin gene family can nonetheless be discerned, and they raise several important and interrelated issues. These range from the developmental problem of actin gene regulation to the evolutionary problem of the origin of the sea urchin actin gene set.

MATERIALS AND METHODS

The methods used in our laboratories for construction and screening of genomic libraries, preparation of recombinant DNAs, ribonucleic acid (RNA) and DNA gel blots, and electron microscope visualization of heteroduplexes have been described, and the procedures can be found in recent papers by Anderson et al. (1), Scheller et al. (R. H. Scheller, D. M. Anderson, J. W. Posakony, L. B. McAllister, R. J. Britten, and E. H. Davidson, *J. Mol. Biol.*, in press), Posakony et al. (J. W. Posakony, R. H. Scheller, D. M. Anderson, R. J. Britten, and E. H. Davidson, *J. Mol. Biol.*, in press), Crain et al. (in press), and Durica et al. (11). The restriction maps shown in Fig. 5 and 6 were obtained by conventional methods, using a combination of partial and double digests. Additional mapping data were derived from electron microscope analysis of heteroduplexes, as described in the text. In the following sections we provide details relevant to particular experiments presented in Results.

DNA sequences. The DNA sequences shown in Fig. 1 and 4 were determined by the Maxam and Gilbert procedure (27), using strand-separated or asymmetrically 5^{32}P -labeled restriction enzyme fragments. Polyacrylamide gels (40 cm; 25, 20, and 8%) were used to display the cleavage products. The restriction fragments used to obtain the sequences shown in Fig. 1 and in Fig. 4 were as follows. The plasmid cDNA clone SpG2 was sequenced in both directions from the *HincII*, *HaeIII*, and *HinfI* sites present in the insert and also from the vector pBR322 *HaeIII* site immediately to the left of the insert (see restriction maps in Fig. 5). The $\lambda\text{SpG2-8}$ sequence shown in Fig. 4 was obtained by first isolating the 4.0-kilobase (kb) *EcoRI* fragment containing actin gene G (see Fig. 6) and then digesting this with *HaeIII* (see lower portion of Fig. 5). Proximal regions of the two fragments separated by the *HaeIII* site at the 3' end of the coding region of the gene were sequenced.

RNA gel blots. The RNA gel blots shown in Fig. 8 were carried out according to B. Seed and D. Goldberg (unpublished data). Total gastrula and oocyte polyadenylated [poly(A)] RNAs were denatured in a mixture of 50% formamide, 2.2 M formaldehyde, and gel buffer by heating at 55°C for 10 min. Tracking dyes were added, and the samples were run on a denaturing 1% agarose gel system in a buffer containing 20 mM sodium morpholinopropanesulfonic acid, 5

mM sodium acetate, 1 mM sodium ethylenediaminetetraacetate (EDTA), and 2.2 M formaldehyde. After electrophoresis the gels were equilibrated with 20× SSC (1× SSC = 0.15 M NaCl–0.015 M sodium citrate, pH 7.5) and transferred by the standard blotting procedure (32). The transfer was continued for about 10 h. Sea urchin ribosomal RNAs, which are about 4,000 and 1,840 nucleotides (nt) long (24), and *Escherichia coli* ribosomal RNAs (2, 3) were used as size standards. After baking, the filters were prehybridized for 2 h in 1× Denhardt solution (10), 50% formamide, 5× SET [1× SET = 0.15 M NaCl, 0.03 M tris(hydroxymethyl)aminomethane-hydrochloride (pH 8), 2 mM EDTA], 0.1% sodium pyrophosphate, and 50 μg of denatured calf thymus DNA carrier per ml. Labeled probes were denatured in boiling water, and the hybridizations were carried out at 37°C for 15 h. In different experiments, 1 to 5 μg of poly(A) RNA was used per lane, and 5×10^5 to 5×10^6 cpm of tracer was present in the hybridization mixtures. Hybridization was carried out in the same solution as the prehybridization, and about 3 ml of hybridization solution was used per lane. Filters were washed in 1× SET at 68°C for 3 to 4 h and then dried and autoradiographed.

Electron microscopy of DNA and RNA-DNA heteroduplexes. The DNA heteroduplexes shown in Fig. 3 and 7 were formed and spread for electron microscopy essentially according to Davis et al. (9). Whole phage or restriction enzyme fragments were mixed in an equimolar ratio and denatured in 0.1 M NaOH–5 mM EDTA at room temperature for 5 min. Tris(hydroxymethyl)aminomethane-hydrochloride was used to neutralize the solution, and formamide was then added to 50%, unless otherwise indicated. These mixtures were allowed to incubate at room temperature for periods such that 10 to 20% of the DNA reacted. R-loop heteroduplexes, such as shown in Fig. 3, were formed according to the method of Kaback et al. (18). Thus, 1 μg of 4,5,8-trimethylpsoralen per ml was added to a 50- $\mu\text{g}/\text{ml}$ DNA solution in 10 mM tris(hydroxymethyl)aminomethane buffer. Samples were irradiated for different lengths of time from 1 to 5 min, and the cross-linking frequency was checked. Samples were used that had about one cross-link per 5,000 base pairs. Cross-linked whole phage DNA (0.1 μg) was mixed with 10 μg of cytoplasmic poly(A) RNA and incubated in a hybridization buffer consisting of 70% formamide–0.4 M NaCl–0.1 M PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid)]–0.01 M EDTA (pH 7.2). The hybridization mixtures were allowed to react for 24 h in a sealed 100- μl capillary tube starting at 55°C. The temperature was decreased by 1°C every 2 h. The mixture was then cooled to 0°C, glyoxal was added to 1.0 M, and incubation was continued at 10°C for an additional 2 h. This mixture was then passed over a 0.4-ml Sepharose 2-B column in 0.5 M NaCl–0.1 M PIPES (pH 7.2)–0.01 M EDTA at 4°C. One-drop fractions were collected, and the excluded fractions were spread for visualization by electron microscopy. Multiple examples of each of the structures shown in the figures were observed. Single-stranded ϕX174 DNA was included as a size standard. A Philips 301 transmission electron microscope was used, and the samples were shadowed with platinum in a Kinney evaporator system.

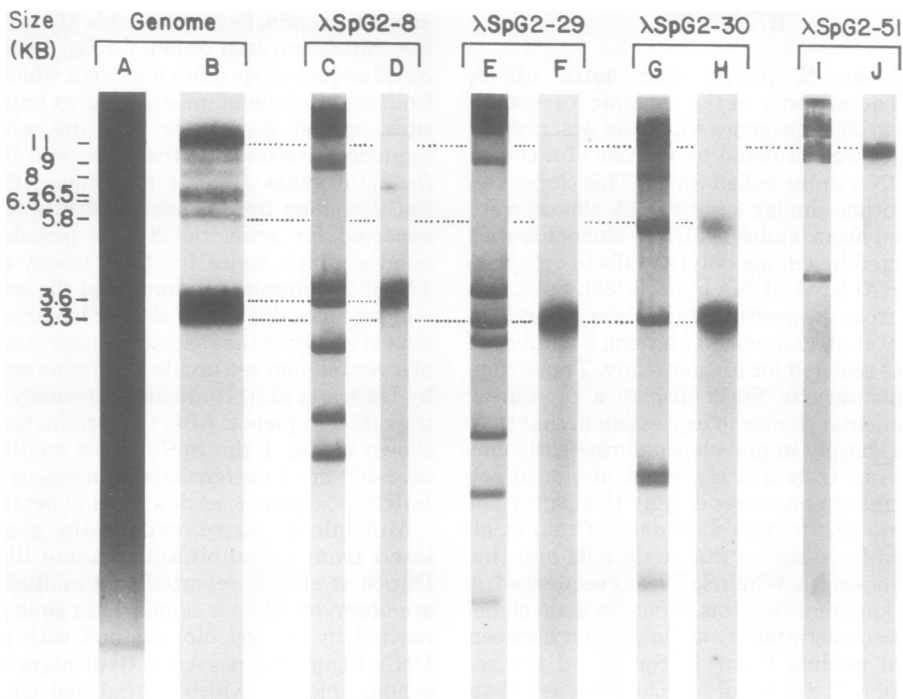


FIG. 2. Gel blot hybridizations with the SpG2 actin probe. The probe was labeled by nick translation, and hybridization was carried out at 68°C in 4× SET. Identical results were obtained whether we used the whole SpG2 clone or a subfragment containing only the coding region. The filters were washed in 1× SET for 3 h at 68°C. Lanes A, C, E, G, and I show the gels stained with ethidium bromide after the DNA was digested with *Eco*RI. Lanes B, D, F, H, and J show autoradiographs of the blotted gels. Lanes A and B display a genome gel blot carried out with DNA from an individual sea urchin. The same DNA was used to construct the λ recombinant genome Library C (1); see text. Sizes (kb) of the actin fragments observed in lane B are indicated in the left margin, measured according to parallel markers (not shown). Blot hybridizations with the λ recombinants indicated were carried out independently with corresponding markers (not shown) in each case. Reactive fragment sizes are: for λ SpG2-8 (lane D), 3.6 and 4.0 kb; for λ SpG2-29 (lane F), 3.3 kb; for λ SpG2-30 (lane H), 3.3 and 5.8 kb; and for λ SpG2-51 (lane J), 11 kb.

digested DNA from various λ recombinants which had been selected in library screens by using SpG2 probes. The three λ recombinants used for the experiments shown in lanes C to H were isolated from a genome library constructed from the DNA of a single sea urchin, and DNA of this same individual was used for the genome blot shown in lanes A and B. This genomic library is known as Library C and was described previously by Anderson et al. (1). It was assembled by the method of Maniatis et al. (26), i.e., ligation of *Eco*RI linkers onto DNA fragments produced by partial digestion with *Hae*III. The *Eco*RI sites which terminate the inserts from Library C do not usually exist in the genome, whereas the internal *Eco*RI sites do. It follows that actin genes which are present on such internal *Eco*RI fragments can be matched by the size of these fragments to those observed in the genome blot, whereas other evidence is required for actin genes located on terminal fragments.

Figure 2 shows that λ SpG2-29 (lane F) and λ SpG2-30 (lane H) both contain the 3.3-kb actin DNA fragment seen in the genome blot. λ SpG2-30 also contains the 5.8-kb actin DNA band observed in the genome (lane H), and λ SpG2-8 (lane D) contains the 3.6-kb actin genome band. The second actin band in λ SpG2-8, at about 4.0 kb, is inferred to be on a terminal fragment since no band of this size appears in the genome blot. The last two lanes of Fig. 2 (I and J) display a blot carried out with λ SpG2-51, which had been isolated from Library A. This is an *Eco*RI partial library (1), constructed from the DNA of a different individual sea urchin. The 11-kb actin DNA fragment in this recombinant also matches a band observed in the genome blot. The recombinants used for the experiment shown in Fig. 2 thus include representatives of four of the eight types of *Eco*RI fragment which contained actin genes in this genome and, in addition, one other actin gene which cannot be assigned to a genome

band from this experiment. Two of the four λ recombinants in the experiment apparently could contain more than one actin gene.

The experiments shown in Fig. 2 are examples of a large number of similar analyses, carried out with several different restriction enzymes and including many different λ recombinants. Library C was screened several times, and over 20 diverse actin isolates were identified and characterized. Some of these were nearly identical to others except for slightly different termini and were not further considered. The following discussion of actin gene organization is based on 10 different isolates from Library C, including the three shown in Fig. 2 (lanes C to H), plus two isolates from Library A, including that shown in Fig. 2 (lanes I and J). Two additional actin clones were isolated from a genomic plasmid library constructed from terminally digested *Hind*III DNA fragments. A partial description of these clones, SpG17 and SpG28, was presented earlier by Durica et al. (11).

Structure of actin genes and identification of five actin gene types. To position the actin gene regions in the λ recombinants, and to determine the approximate location of their intervening sequences, R-loops were formed with embryo poly(A) RNA and visualized in the electron microscope. Regions of nonhomology between different actin genes were also located by reacting appropriate restriction enzyme fragments from the λ isolates and examining the heteroduplexes for single-stranded segments. A sample of these observations is reproduced in Fig. 3.

Figure 3A displays an R-loop structure consisting of a gastrula cytoplasmic poly(A) RNA molecule and the DNA of λ SpG2-29. As shown above (Fig. 2, lane F), this isolate contains a 3.3-kb *Eco*RI fragment which reacts with the SpG2 actin probe. Two intervening sequences are evident in the R-loop. One of these is known to occur at position 121/122 in the amino acid sequence (11), and the second occurs at position 203/204 in the sequence (Crain et al., unpublished data). These intervening sequences turn out to provide useful markers for particular actin genes, since in certain cases they are sufficiently divergent to give rise to single-stranded bubbles when different cloned genes are reacted with each other and spread at high criterion (Fig. 3B and C). The two actin *Eco*RI fragments from λ SpG2-8 (see Fig. 2, lane D) were isolated and hybridized with each other in 50% formamide (Fig. 3B) and in 65% formamide (Fig. 3C). The heteroduplexes show that, as implied in Fig. 2, λ SpG2-8 does indeed contain two complete actin genes. The more 5' of the intervening sequences in these genes remain single stranded under both

criterion conditions, whereas both intervening sequences remain single stranded at the higher criterion. Very approximately, this indicates that the 5' intervening sequences of these two linked actin genes must be at least 25% divergent from each other, whereas the 3' intervening sequences must be at least 15% divergent. To the 5' side of the coding region the two λ SpG2-8 actin genes appear completely nonhomologous. However, much of the 3' flanking region remains in duplex form, even at the higher criterion.

The heteroduplexes shown in Fig. 3B and C clearly establish two distinct actin gene types. The existence of a third type is evident in Fig. 3D, in which a heteroduplex between one of the λ SpG2-8 actin genes and the λ SpG2-29 actin gene is shown. Although the intervening sequences of this pair of genes are at least relatively homologous, the 3' flanking regions immediately abutting the coding portions of the gene are nonhomologous. The presence of diverse sequences at the 3' ends of these actin genes provides a means of determining the affinity of the cDNA clone SpG2, since as noted above this clone contains several hundred nucleotides of 3' untranslated sequence. In Fig. 3E a heteroduplex between the same λ SpG2-8 gene as used in Fig. 3D (i.e., that carried on the 4.0-kb *Eco*RI fragment) and SpG2 itself is shown. It can be seen that these two sequences diverge completely about a third of the way along the SpG2 insert, i.e., somewhere near the position of the translation stop codon. It follows that the transcript which is represented in clone SpG2 could not have derived from an actin gene of either of the two types present in λ SpG2-8. Data shown below demonstrate that the gene carried in λ SpG2-29 is of the same type as that repre-

To determine the exact position of sequence divergence between this cDNA clone and the λ SpG2-8 gene in the heteroduplex of Fig. 3D, the sequences flanking the coding regions on the 3' side were obtained. These sequences are compared in Fig. 4. Here it can be seen that an abrupt point of divergence occurs just 13 nt beyond the translation stop codon in SpG2. The sequence of this cDNA clone then continues on for an additional 427 nt. This region of SpG2 is notably AT-rich (62.5% AT), compared to the coding region (47.4% AT), and it contains numerous translation stop codons in all three reading frames.

To distinguish the various actin genes further, restriction maps of the gene regions were assembled and are compared in Fig. 5. The uppermost map represents the plasmid clone SpG28 and shows clearly that the gene carried on this clone represents a fourth actin gene type. This gene has two additional intervening sequences, and

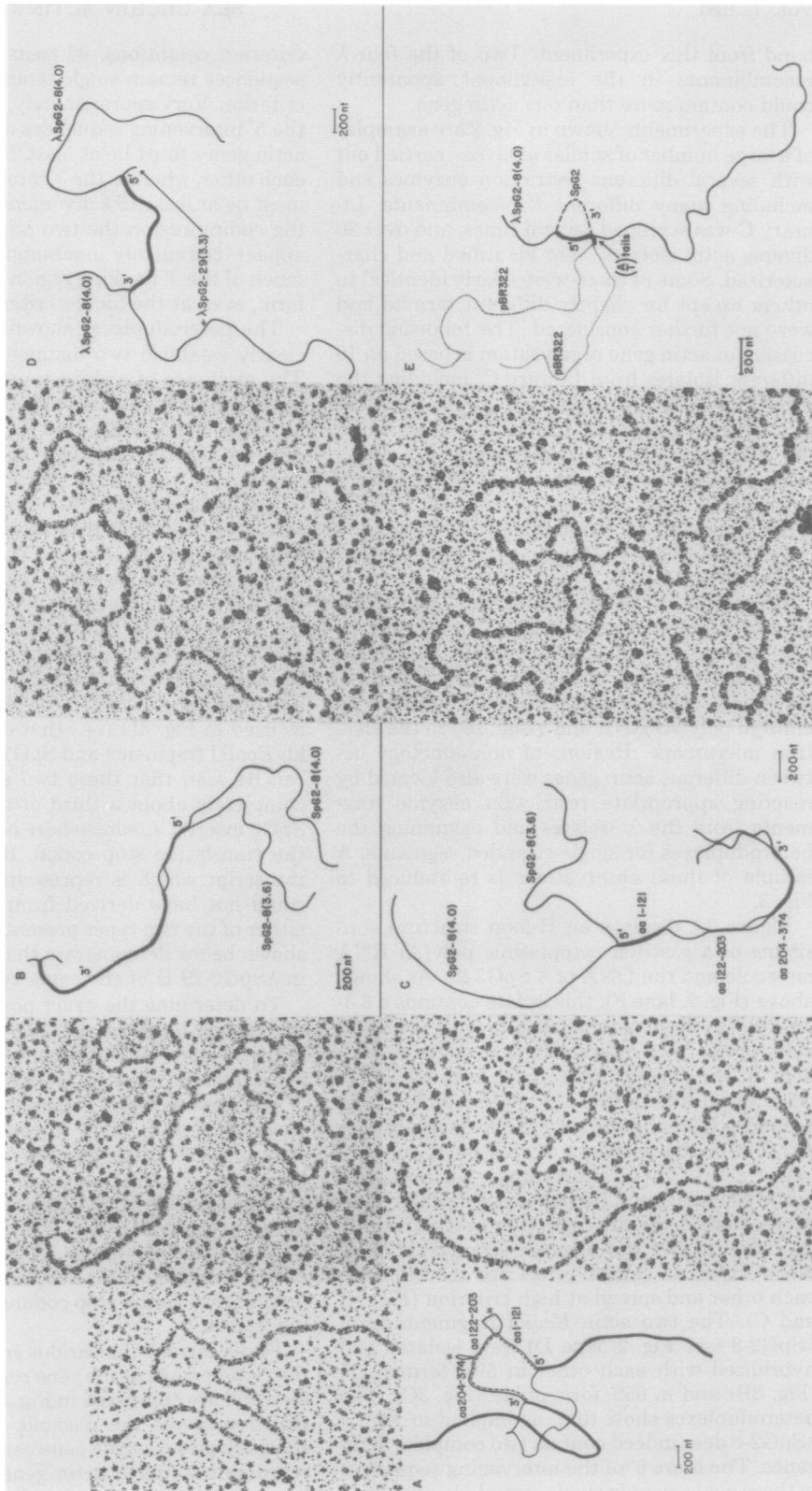


Fig. 3. R-loop and DNA heteroduplexes involving diverse actin genes. (A) R-loop structure formed by reacting poly(A) RNA from sea urchin gastrula with the DNA of λ SpG2-29. To improve stability the DNA was cross-linked with a psoralen derivative before hybridization as described in the text. On the interpretive tracing the protein coding segments are designated according to the amino acid sequence of rabbit skeletal muscle actin, which is highly homologous to sea urchin actin (11; Crain et al., unpublished data; Fig. 1 of this paper). Positions of the intervening sequences are derived from Durica et al. (11), who determined them for clone SpG17. As shown in Fig. 5, this actin gene and that carried in λ SpG2-29 are identical. Polarity assignments were derived from sequences shown in Fig. 1 and restriction map overlaps. Dashed line, RNA; thin line, single-stranded DNA; thick line, double-stranded DNA. Asterisks indicate the intervening sequence loops that had been spliced out of the mature poly(A) RNA. About 50% of the structures observed displayed both R-loops. (B) and (C) DNA of λ SpG2-8 was digested with *EcoRI*, and the 3.6-kb and 4.0-kb fragments which react with the SpG2 probe (Fig. 2) were isolated by gel electrophoresis. These fragments were reacted with each other at room temperature in a medium containing (B) 50% formamide or (C) 65% formamide (see the text). They were then spread for electron microscopy. Lines on the interpretive tracings are as in (A). 3'. Approximate 3' end of the coding region of the gene. (D) Heteroduplex between the 4.0-kb *EcoRI* fragment of λ SpG2-8 and the 3.3-kb *EcoRI* fragment of λ SpG2-29, reacted in the 50% formamide medium as in (B). Lines on the interpretive tracing are as in (A). (E) Heteroduplex between the same 4.0-kb *EcoRI* fragment of λ SpG2-8 as used in (D) and the DNA of the actin cDNA clone SpG2; the reaction was carried out in the 50% formamide medium. SpG2 DNA was first linearized by digestion with *PvuII*. The two poly(A) tails of the cDNA clone have reacted to form a stemmed loop within which lies the actin gene insert. This insert is homologous for about $\frac{1}{2}$ of its length with the 3' coding region of the λ SpG2-8 actin gene. The 3' untranslated regions are not homologous. In the interpretive tracing the dotted lines represent the 3' arms of SpG2, terminating in the *PvuII* site; the dashed line represents the single-stranded 3' region of the SpG2 insert; the heavy line represents the double-stranded heteroduplex; and the light line represents single-stranded DNA of the λ SpG2-8 fragment.

almost all of the restriction sites where comparisons are possible are unique to SpG28. According to the data of Crain and his associates (to be presented elsewhere), the SpG28 intervening sequences occur at positions 41/42, 121/122, 203/204, and 267/268 of the amino acid sequence. Figure 5 also shows that the plasmid clone SpG17 is homologous at every mapped site with the actin gene carried in λ SpG2-29, i.e., with that species of actin gene of which the 3' coding region is carried on a 3.3-kb *EcoRI* band (cf. Fig. 2). The regions directly in and around the actin genes are not polymorphic, in that genome blots carried out with the SpG2 probe on DNA from seven different individual sea urchins are almost identical. That is, 55 of 56 bands observed were in identical positions (not shown). Thus, even though they derive from different sea urchin genomes, SpG17 and λ SpG2-29 display the same 3.3-kb *EcoRI* actin DNA fragment.

As expected from the foregoing data, Fig. 5 also shows that the two actin genes carried on λ SpG2-8 are different in 5' flanking map sites from each other as well as from the gene carried on SpG17. A fifth gene type is also present in our collection, though so far little is known about it, and it is not included in Fig. 5. This is the actin gene carried on the 11-kb band of λ SpG2-51 (Fig. 2, lane J). There is no known restriction site homology between this gene and any of the others outside of the coding region.

In the lower portion of Fig. 5 the distinctive 3' untranslated region of SpG2 (Fig. 4) is compared with those of two other cDNA clones and with the equivalent regions of several of the cloned genes present on our λ isolates. The first three maps show that the cDNA clones are very similar in that they all contain a 3' untranslated sequence of almost exactly the same length, as well as various amounts of coding sequence. The 3' noncoding regions of all of these clones contain a pair of restriction sites (*HpaII* and *HinfI*) in precisely the same central location. However, they differ in their *HincII* sites. SpG2 and SpP57 lack the central *HincII* site in the tail region which is present in SpP110. On the other hand, the terminal *HincII* site which precedes the 3' poly(A) tract in SpG2 by 1 nt is absent from the other two cDNA clones. These differences almost certainly reflect sequence differences in the genomic actin genes. Thus the succeeding maps in Fig. 5 show that the actin genes found in λ SpG2-28, λ SpG2-16, and λ SpG2-29 all lack the central *HincII* site found in the cDNA clones, but include the terminal one, whereas the λ SpG2-52 actin genes lack the terminal *HincII* site and include the central one. We conclude that the transcript species represented in SpG2 derived from a gene of the type carried on a 3.3-



FIG. 4. Divergent sequences of 3' untranslated regions from different actin genes. The 4.0-kb *EcoRI* fragment of λ SpG2-8 was cut with *HaeIII*, and the sequence of the region immediately surrounding this site at the 3' end of the coding region of the gene was determined as described in the text. This sequence is compared to the equivalent sequence in SpG2 (see Fig. 1 for the coding region sequences of SpG2). The sequences are numbered with respect to the SpG2 insert, i.e., as in Fig. 1. Only the coding strands of both sequences are shown. The asterisk indicates a divergent nt in an otherwise homologous region. As indicated, the SpG2 sequence is shown on top and the λ SpG2-8 sequence is shown below. The coding region sequence carried in this copy of the actin gene is perfectly identical to that in SpG2 (indicated by the dashed line). The two sequences are homologous until SpG2 nt 201, when they suddenly become entirely distinct. The 3' untranslated region of SpG2 is 443 nt long, with the poly(A) tail beginning at nt 628. The A_2UA_3 site commonly observed before the poly(A) tail of eucaryotic mRNA's is not present in this clone (see text).

kb *EcoRI* fragment. The restriction site data suggest that the transcript represented in SpP110 may well derive from a very closely related gene of the type carried on λ SpG2-52, which in the genome occurs on an 11.5-kb *EcoRI* fragment. In any case, however, all three of the cDNA clones and all of these four λ isolates carry actin genes which possess more or less the same distinctive 3' untranslated tail sequence. The characteristic of this actin gene type is the presence of the closely apposed *HinFI* and *HpaII*

sites in the middle of this tail region, which as shown above lacks all homology with the equivalent regions of the two λ SpG2-8 actin gene types.

An experiment was also carried out with a probe consisting of the SpG2 sequence element bounded by the 3' terminal *HincII* site and the nearby *HpaII* site (sites 12 and 7 in Fig. 5). This probe contains no coding sequence and therefore is specific for the 3' untranslated sequence of the message represented by SpG2. The probe was

reacted by blot hybridization with λ SpG2-16, -51, and -52 and SpG28. As expected, a strong reaction was observed with the λ SpG2-16 and -52 actin genes, and no reaction whatsoever was detected with the λ SpG2-51 actin gene. The

probe also reacted with the plasmid clone SpG28, despite the basic difference in structure between this gene and the other actin gene types studied. Thus, the 3' tail of the SpG2 sequence shown in Fig. 4 is shared by members of only

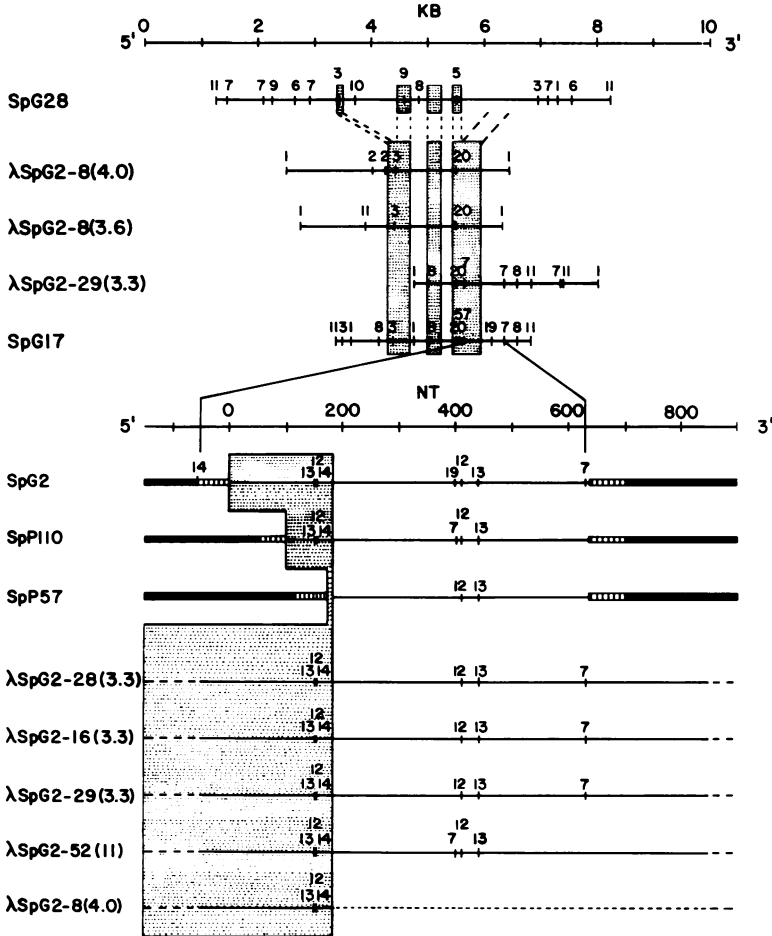


FIG. 5. Restriction map comparisons of various actin gene isolates. Shaded regions indicate the position of nucleotides coding for actin proteins. The positions of the coding regions are based on restriction enzyme mapping, electron microscopy of R-loops and DNA heteroduplexes such as shown in Fig. 3, and the direct nucleotide sequence data obtained by Durica et al. (11) and Crain et al. (in press). Intervening sequences in the plasmid clone SpG28 are located at amino acid positions 41/42, 121/122, 203/204, and 267/268. The positions of the intervening sequences in the λ SpG2 isolates shown and in SpG17 are at amino acids 121/122 and 203/204. Restriction enzyme designations: 1, EcoRI; 2, SalI; 3, BamHI; 4, SacI; 5, XhoI; 6, XbaI; 7, HincII; 8, PstI; 9, BglII; 10, ClaI; 11, HindIII; 12, HpaII; 13, HinfI; 14, HaeIII; 15, HhaI; 16, SmaI; 17, AvaI; 18, KpnI; 19, BstEII; 20, PvuII. Mapping has been done as follows: SpG28 with enzymes 1, 3, 5, 6, 7, 8, 9, 10, and 11; the λ SpG2-8 (4.0) fragment with enzymes 1, 2, 3, 11, and 20; the λ SpG2-8 (3.6) fragment with enzymes 1, 2, 3, 11, and 20; the λ SpG2-29 (3.3) fragment with enzymes 1, 2, 3, 7, 8, 11, and 20; and SpG17 with enzymes 1, 2, 3, 5, 7, 8, 11, 19, and 20. The bottom part of the figure shows an expanded view of the 3' regions of five actin genes compared with three cDNA clones, with several additional restriction sites, as shown. cDNA clone nomenclature (22) indicates the origin of the clone in a gastrula (G) library or a pluteus (P) library. The scale in nucleotides for this part of the figure is numbered such that base pair 1 is the 5' end of the SpG2 cDNA insert as in Fig. 1 and 4. The thick solid lines represent the pBR322 vector of the cDNA clones, and the crossed boxes represent their poly(A) tails. The dashed line in the 3' tail region of the actin gene carried on the 4.0-kb EcoRI fragment of λ SpG2-8 signifies the complete lack of homology with the equivalent region of any of the other actin genes shown in this part of the figure (see Fig. 4).

two of the five actin gene types, namely, that type carried on the 3.3-kb and 11.5-kb *EcoRI* fragments in the genome and in SpG28. The same sequence is represented in all three of the actin cDNA clones.

We are now in a position to consider the overall organization of the actin gene family. As described below, some insight into this can be obtained from a comparison of flanking restriction sites in the actin isolates derived from Li-

brary C, taking into account the diverse actin gene types thus far demonstrated.

Linkage of nonallelic actin genes. When restriction maps of the inserts in the actin λ recombinants were compared, it became apparent that there are several nonallelic copies of at least some of the actin gene types in the genome (Fig. 6). The 10 actin recombinants whose maps appear above the horizontal line in Fig. 6 are library C isolates. Seven recombinants which

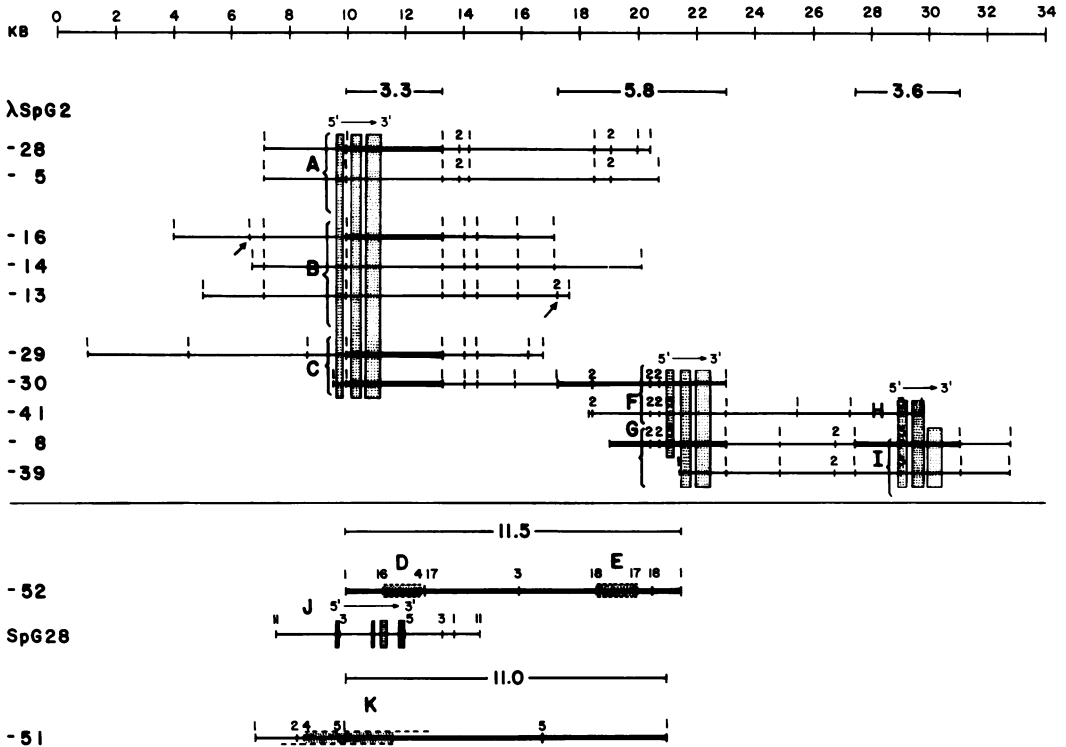


FIG. 6. Partially linked arrangement of actin genes and identification of nonallelic actin gene copies. Restriction maps of the 10 Library C actin isolates discussed in the text are shown above the horizontal line. The individual recombinants are identified along the left-hand margin, and a scale in kb is shown along the top. Actin gene regions coding for protein sequence are shaded as in Fig. 5. *EcoRI* fragments designated by heavy lines are those discussed earlier in the context of other experiments (Fig. 2, 3, and 5). These are, respectively, λ SpG2-28, -16, and -30, the 3.3-kb fragments; λ SpG2-30, the 5.8-kb fragment; λ SpG2-8, the 4.0-kb and 3.6-kb fragments; λ SpG2-52, the 11.5-kb fragment; λ SpG2-51, the 11.0-kb fragment. The bars above the maps indicate the actin gene types 3.3, 5.8, and 3.6, distinguished by the size of the *EcoRI* fragments in genomic DNA bearing actin genes which react with the 3' actin gene probe SpG2. Similar bars below the horizontal line denote the 11.5- and 11-kb *EcoRI* fragments bearing actin genes. Terminal *EcoRI* sites on the λ recombinants are in general expected to be linker sites rather than naturally occurring genomic *EcoRI* sites. Thus for example, as inferred earlier (Fig. 2), the 4.0-kb fragment of λ SpG2-8 is a terminal one. Numbers represent restriction sites, as given in the legend to Fig. 5. Enzymes 1 and 2 were used to map λ SpG2-28, -5, -16, -14, -13, -29, and -30. Enzymes 1, 2, and 3 were used to analyze λ SpG2-41, -8, and -39. Enzymes 3, 4, 16, 17, and 18 were used to map the single *EcoRI* fragment carried in λ SpG2-52; and enzymes, 1, 2, 4, and 5 were used to map the 11.0-kb *EcoRI* actin fragment carried in λ SpG2-51. A partial restriction map of SpG28 is reproduced from Fig. 5, for comparison. Capital letters indicate individual nonallelic actin genes, inferred as described in the text. Brackets enclose genes that could be allelic variants. Arrows mark two restriction sites distinguishing what probably represents the two alleles of actin gene B. The internal structure of actin genes D, E, and K is unknown. These genes are indicated only by dashed lines between restriction sites that limit fragments reacting with the SpG2 probe in gel blot experiments.

contain a 3.3-kb *EcoRI* actin fragment are included; based on the flanking sequence restriction site differences, these fall into three classes, denoted A, B, and C. We believe these classes to represent at least three nonallelic copies of the "3.3" actin gene type. This can be seen as follows. Outside of the *EcoRI* site which forms the right-hand boundary (i.e., the 3' boundary, with respect to gene orientation) of the reactive 3.3-kb fragment, all of the sites in λ SpG2-5 and λ SpG2-28 differ from those in the equivalent portion of any of the other isolates shown. This establishes the A actin gene as a separate entity. The first three, but not the fourth, *EcoRI* sites flanking the actin B and C genes of Fig. 6 to the 3' side are all similar in λ SpG2-16, -14, -13, -29, and -30, but after this the B representatives λ SpG2-14, -16, and -13 differ in every site from the C gene representative λ SpG2-30. Furthermore, the C gene isolate λ SpG2-29 contains 5' *EcoRI* sites absent from the equivalent region of the B gene isolates λ SpG2-16, -14, and -13, and vice versa. It is not possible that these five B and C gene isolates are merely polymorphic variants representing a single allele, because there could exist only two alleles for each gene in the diploid genome of the single individual sea urchin that provided the DNA of Library C. In the case of the B gene, both alleles may have been recovered. Thus two restriction sites vary among the three B gene isolates (arrows) so that λ SpG2-16 and -14 are likely to represent one allele of the B gene while λ SpG2-13 may represent the other. We conclude that there are no less than three nonallelic copies of the 3.3 actin gene type, namely, A, B, and C. This may account in part for the extra intensity of the 3.3-kb *EcoRI* band in the genome blot of Fig. 2. It will be recalled that A, B, and C all share the distinctive 3' restriction pattern flanking the coding region (Fig. 5), and, where tested, they form perfect heteroduplexes in the gene region. There could of course be more than three copies of the 3.3 actin gene type if the variants just discussed were actually separate genes rather than polymorphic allelic partners, or if there were other 3.3 type genes which we failed to isolate. In addition, as discussed above, the two actin genes indicated on λ SpG2-52 in Fig. 6 are very closely related to the 3.3 type genes A, B, and C, since they contain the same characteristic 3' untranslated tail sequence. These genes, labeled D and E in Fig. 6, are not yet characterized, and their structure is unknown, except that surrounding restriction sites distinguish the environment of D from that of E. Provisionally we consider them the same actin gene type as genes A, B, and C, though further study might of course

reveal distinctive features.

Similar arguments lead to the conclusion that there are at least two nonallelic copies of the "5.8" type of actin gene, F and G, and two copies of the "3.6" gene type, H and I (Fig. 6). As discussed above, these numerical designations of actin gene type refer to the length of the *EcoRI* band containing these actin genes and recognized in genome blots and in our λ recombinants. We define gene F as the second gene on λ SpG2-30. This recombinant contains both a 3.3 and a 5.8 type of actin gene, separated by about 9 kb of other sequence. A gene almost identical in mapped sites to this 5.8 gene is present on the left end of λ SpG2-41 (the *BamHI* site present in λ SpG2-41 and -8 in this gene could well be present in the copy of the gene carried in λ SpG2-30, since this enzyme was not used to map the latter isolate). Gene F is in turn linked to a third gene of the 3.6 type, labeled H in Fig. 6. A similar pair of genes is present on λ SpG2-8, namely, genes G and I. The latter are clearly separate, nonallelic copies of the 5.8 and 3.6 gene types, since the DNA separating G and I displays different restriction sites from the DNA separating F and H.

A further verification of the linked arrangement of the diverse actin gene types on λ SpG2-8 and λ SpG2-30 is shown in Fig. 7. This figure displays heteroduplexes formed by reacting these two isolates. Two forms of such heteroduplexes exist. In the simpler form seen in Fig. 7A, only the 5.8 type genes F and G have hybridized, along with the overlapping 5' sequence (cf. maps in Fig. 6). In the more complex form shown in Fig. 7B, gene C of λ SpG2-30 has reacted with gene I of λ SpG2-8 as well. The small intervening sequence loop-out seen in Fig. 7A suggests but does not prove that the λ SpG2-8 gene G is indeed the nonallelic close relative of gene F, the arrangement indicated in Fig. 6, i.e., that it is gene H which is linked to gene C via gene F in the genome. However, this is not unequivocally demonstrated by the present evidence. It is possible that there are two nonallelic genes included in the bracket labeled F in Fig. 6, which might be termed F and F'. Thus, we cannot exclude the possibility that C-F and F'-H linkages all occur in separate genomic locations.

Summarizing the main consequences of the overlap patterns shown in Fig. 6, we conclude that there are at least three 3.3 type actin genes, A, B, and C; two additional very closely related genes, D and E; at least two 5.8 type actin genes, F and G; and at least two 3.6 type actin genes, H and I. Considering in addition the two actin gene types carried on the SpG28 clone and on λ SpG2-

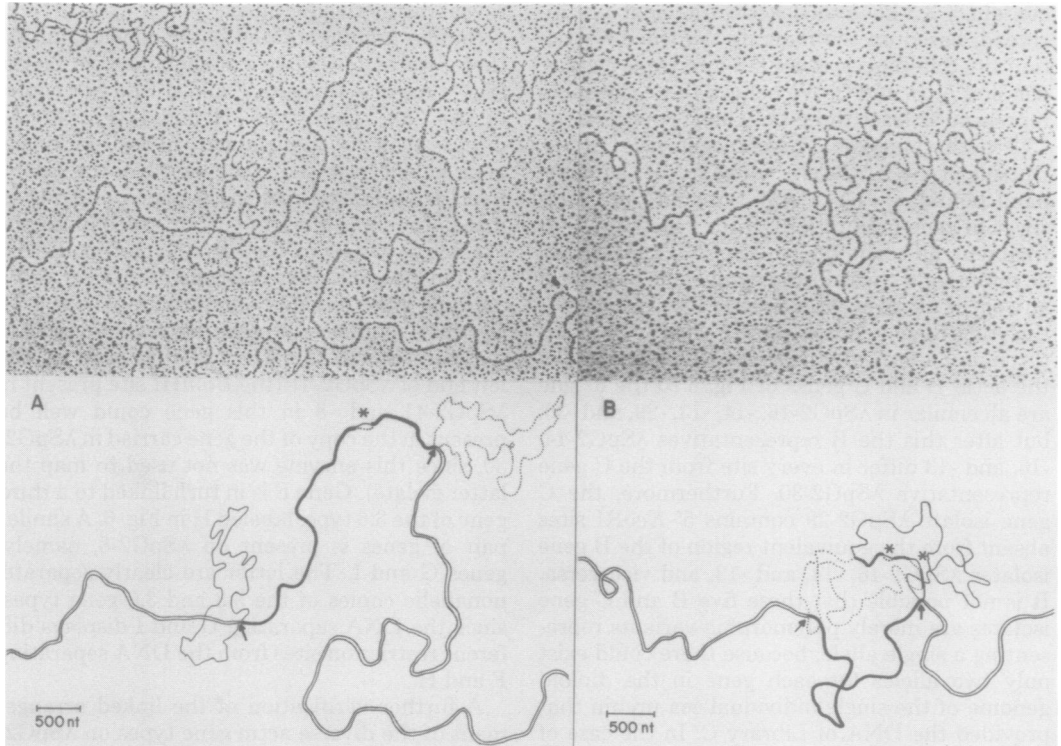


FIG. 7. Heteroduplexes between λ SpG2-8 and λ SpG2-30. Dotted lines in the tracings, single-stranded λ SpG2-8 DNA; thin solid lines, single-stranded λ SpG2-30 DNA; thick solid lines, λ SpG2-8/ λ SpG2-30 double-stranded duplex; double lines, the Charon 4 vector. Arrows point to the junctions between the vector arms and the *S. purpuratus* DNA. Asterisks indicate positions of single-stranded intervening sequences. A 60% formamide criterion was applied, as described in the text. (A) Heteroduplex in which the region overlapping between the two λ recombinants has hybridized to form a linear duplex structure interrupted only by a single-stranded intervening sequence (see text and maps in Fig. 6). The hybridized region includes the 5.8 type actin genes F and G. The remainder of both inserts is present as single-stranded loops. (Figure 6 shows that these loops contain additional copies of actin genes, namely, gene C in the λ SpG2-30 loop and gene I in the λ SpG2-8 loop.) (B) Here this second pair of actin genes has also hybridized, leaving only the nonhomologous flanking sequences single stranded. The intervening sequences are present as single-stranded bubbles in the gene C/gene I heteroduplex, as expected (see Fig. 3) since gene C is a 3.3 type gene and gene I is a 3.6 type gene.

51 (Fig. 6, labeled J and K, respectively), this analysis requires a minimum of 11 nonallelic actin genes and 5 actin gene types in the *S. purpuratus* genome. Linkage of actin genes is apparently common, and in all cases, with the possible exception of genes D and E, the linked genes are of different rather than similar type. It is probable that at least one 30-kb region of the *S. purpuratus* genome contains three linked, similarly oriented actin genes; a 3.3 type gene, C; a 5.8 type gene, F; and a 3.6 type gene, H. In addition, at least one other 12-kb region exists which contains both a 5.8 type gene, G, and a 3.6 type gene, I. There is as yet no evidence regarding the relative location of actin genes not so far found to be physically linked to other actin genes, that is, genes A, B, J, and K.

The embryonic actin gene transcripts.

The restriction map and sequence analyses shown in Fig. 4 to 6 establish that our actin cDNA clones represent transcripts which were derived from some or all of the closely related genes A to E. We next consider the problem of positioning the two size classes of embryo transcripts on our maps of the cDNA clones and the genomic actin genes. Several relevant RNA gel blot experiments are shown in Fig. 8. Lane A demonstrates the 1.8- and 2.2-kb actin mRNA transcripts present in gastrula poly(A) RNA as reported earlier by Crain et al. (in press). These transcripts are here visualized by reaction with a 5' coding region probe (labeled I in Fig. 8) derived from the 3.3 type gene C. A 3' coding region probe (II) from the same type of gene also reacts strongly with both of these transcripts, as shown in lane B. However, probe III,

also isolated from a 3.3 type gene, reacted significantly only with the longer transcript. This can be seen in lane C of Fig. 8. Probe III does not overlap by more than a few nucleotides with any of the cDNA clones, since as illustrated in Fig. 4 and 5 the cDNA clones terminate at or very near the *HincII* site which constitutes the left boundary of probe III. This experiment shows clearly that the longer actin transcript derives from 3.3 type genes, i.e., genes A to C, or

from other very closely related genes which may share this 3' sequence element (e.g., genes D, E, or J). In the maternal RNA of the unfertilized egg, by far the predominant form of transcript observed is the 2.2-kb message (Crain et al., in press; Fig. 8, lane D of this paper).

To determine which transcript class is represented by the cDNA clones, the probe labeled IV in Fig. 8 was reacted with gastrula and egg poly(A) RNA (lanes E and F). This probe re-

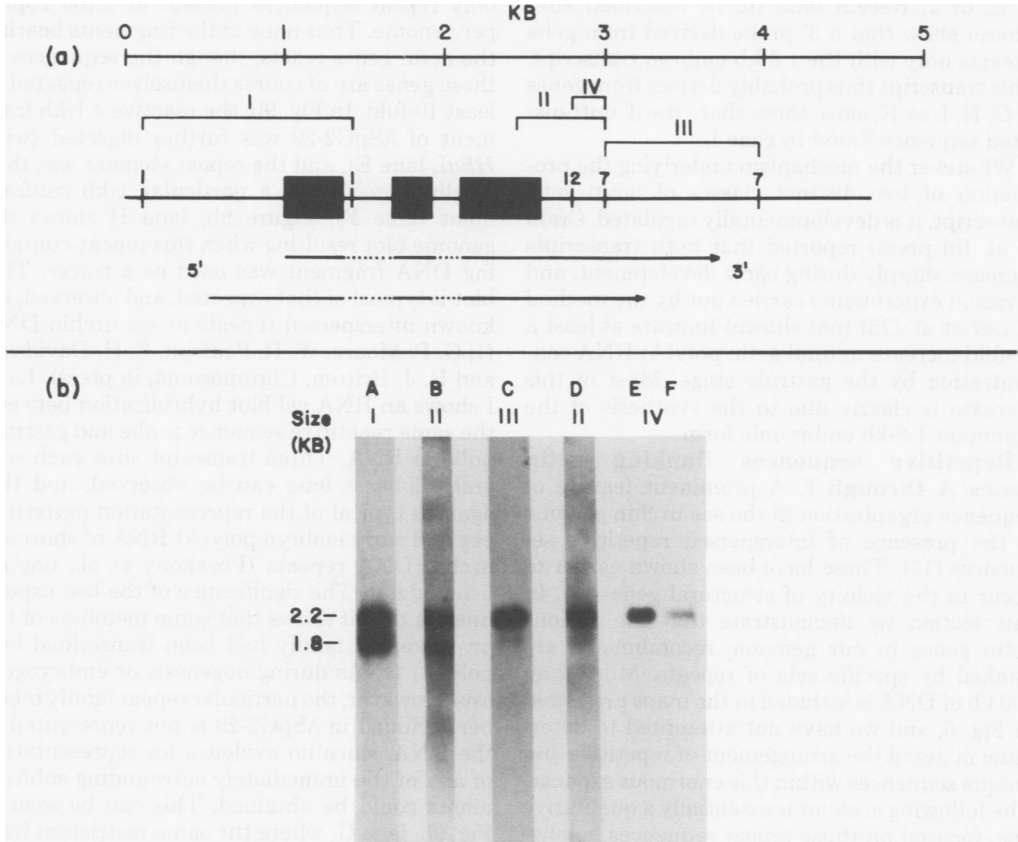


FIG. 8. Reactions of cloned DNA fragments with embryo actin mRNA's. (a) Map of the relevant region of the 3.3 type of actin gene, reproduced from data presented in Fig. 5 and 6. A length scale is shown at the top. Numbers represent restriction sites: 1, *EcoRI*; 7, *HincII*; 12, *HpaII*. Probes used in the RNA gel blot hybridizations of (b) are identified by brackets. Probe I is the 1.3-kb *EcoRI* fragment from λ SpG2-29 which contains the 5' coding region of the actin gene (Fig. 6). Probe II is the SpG2 cDNA clone (Fig. 5). Probe III is the 0.95-kb subfragment released by *HincII* digestion of the 3.3-kb *EcoRI* fragment containing the major portion of the actin gene in λ SpG2-29 (Fig. 5). There was a very slight contamination of this probe with a coding region fragment, accounting for the faint second band seen in lane C (b). Probe IV was prepared by digesting SpG2 with *HincII*, labeling by the kinase reaction, and then digesting with *HpaII*. The appropriate labeled fragment was isolated from a gel for use in the RNA blot experiments. The arrows show the extreme positions on the generalized map of the 3.3 type of actin gene that could be occupied by the 2.2-kb transcript. The upper arrow indicates the transcript position if the message begins at the 5' end of the coding sequence, and the lower arrow indicates the location if it ends about 200 nt beyond the *HincII* site which forms the left-hand boundary of probe III. (b) RNA gel blot hybridization experiments with total gastrula poly(A) RNA (lanes A, B, C, and E) and unfertilized egg poly(A) RNA (lanes D and F). Lanes A to C contained 1 μ g of poly(A) RNA, and lanes D to F contained 5 μ g. The RNA was blotted onto nitrocellulose paper and reacted with the indicated probe fragments as described in the text.

acted only with the 2.2-kb message. The experiment illustrates that the reactive gastrula transcript is the same size as the predominant egg transcript shown in lanes D and F. Therefore the cDNA clones also represent the long transcripts. It follows that reverse transcriptase initiated synthesis of the cloned cDNA at a point other than the actual 3' terminus of the message, which may explain the absence of the A₂UA₃ sequence in SpG2 (Fig. 4). In summary, the 2.2-kb embryo actin transcript derives from genes A to E, or J. Recent data (to be described elsewhere) show that a 3' probe derived from gene I reacts only with the 1.8-kb embryo transcript. This transcript thus probably derives from genes F, G, H, I, or K, since these share the 3' untranslated sequence found in gene I.

Whatever the mechanism underlying the production of two distinct classes of actin gene transcript, it is developmentally regulated. Crain et al. (in press) reported that both transcripts increase sharply during early development, and titration experiments carried out by the method of Lev et al. (25) (not shown) indicate at least a 10-fold increase in total actin poly(A) RNA concentration by the gastrula stage. Most of this increase is clearly due to the synthesis of the dominant 1.8-kb embryonic form.

Repetitive sequences flanking actin genes A through I. A prominent feature of sequence organization in the sea urchin genome is the presence of interspersed repetitive sequences (15). These have been shown earlier to occur in the vicinity of structural genes (8). In this section we demonstrate that the various actin genes in our genomic recombinants are flanked by specific sets of repeats. More than 100 kb of DNA is included in the maps presented in Fig. 6, and we have not attempted to determine in detail the arrangement of repetitive and unique sequences within this enormous expanse. The following account is essentially a qualitative one, focused on those repeat sequences belonging to families which are sufficiently prominent to permit their detection by relatively simple blot hybridization methods. Such sequences were identified by digesting the appropriate actin λ isolates with restriction enzymes and reacting the resulting DNA fragments with genomic DNA labeled by nick translation. The presence of repeats was then verified by isolating and labeling the reactive DNA fragments and using them as probes with blots of restricted genomic DNA. We also attempted to determine whether the repeat families whose members are present in flanking regions of the actin genes are represented in embryo (or egg) RNAs, as has been found previously for a variety of randomly chosen cloned repetitive sequence elements (5, 6).

Some examples of these experiments are reproduced in Fig. 9.

Figure 9 shows two experiments in which repeat sequences were localized in specific *Eco*RI fragments of the λ SpG2-8 and λ SpG2-29 isolates. The reactions carried out with labeled genomic DNA revealed that two fragments of λ SpG2-29 (lane B) and three fragments of λ SpG2-8 (lane D) hybridized easily detectable amounts of tracer. The sensitivity of this experiment is not known, but it was probably sufficient to detect only repeat sequences present at ≥ 100 copies per genome. Thus none of the fragments bearing the actin genes reacts, though the sequences of these genes are of course themselves repeated at least 10-fold. In Fig. 9b, the reactive 4.1-kb fragment of λ SpG2-29 was further digested (with *Hha*I, lane E), and the repeat element was thus localized mainly to a particular 1-kb subfragment (lane F). Figure 9b, lane H shows the genome blot resulting when this repeat-containing DNA fragment was used as a tracer. This blot is typical of that expected, and observed, for known interspersed repeats in sea urchin DNA (1; G. P. Moore, W. R. Pearson, E. H. Davidson, and R. J. Britten, *Chromosoma*, in press). Lane I shows an RNA gel blot hybridization between the same repetitive sequence probe and gastrula poly(A) RNA. Three transcript sizes each several kilobases long can be observed, and this again is typical of the representation pattern in egg and early embryo poly(A) RNA of short sea urchin DNA repeats (Posakony et al., unpublished data). The significance of the last experiment is that it shows that some members of the probe repeat family had been transcribed into poly(A) RNAs during oogenesis or embryogenesis. However, the particular repeat family member included in λ SpG2-29 is not represented in the RNA, since no evidence for representation of any of the immediately surrounding subfragments could be obtained. This can be seen in Fig. 9b, lane G, where the same restriction fragments shown in lane E are reacted with a highly labeled gastrula cDNA tracer. Only the single 1-kb band containing most of the repeat sequence (lane F) reacted. Figure 9c contains a set of experiments similar to those shown in Fig. 9b, but carried out on the 2.7-kb *Eco*RI fragment of λ SpG2-8. According to Fig. 9a, lane D, this fragment also contains a prominent repetitive sequence. All of the repeat elements studied in other experiments displayed a pattern of RNA representation similar to that seen in Fig. 9b and c. That is, they are all represented in multiple transcripts in the poly(A) RNA, but in each case these transcripts almost certainly derived from family members other than those near the actin genes we have studied. In all cases the RNA gel

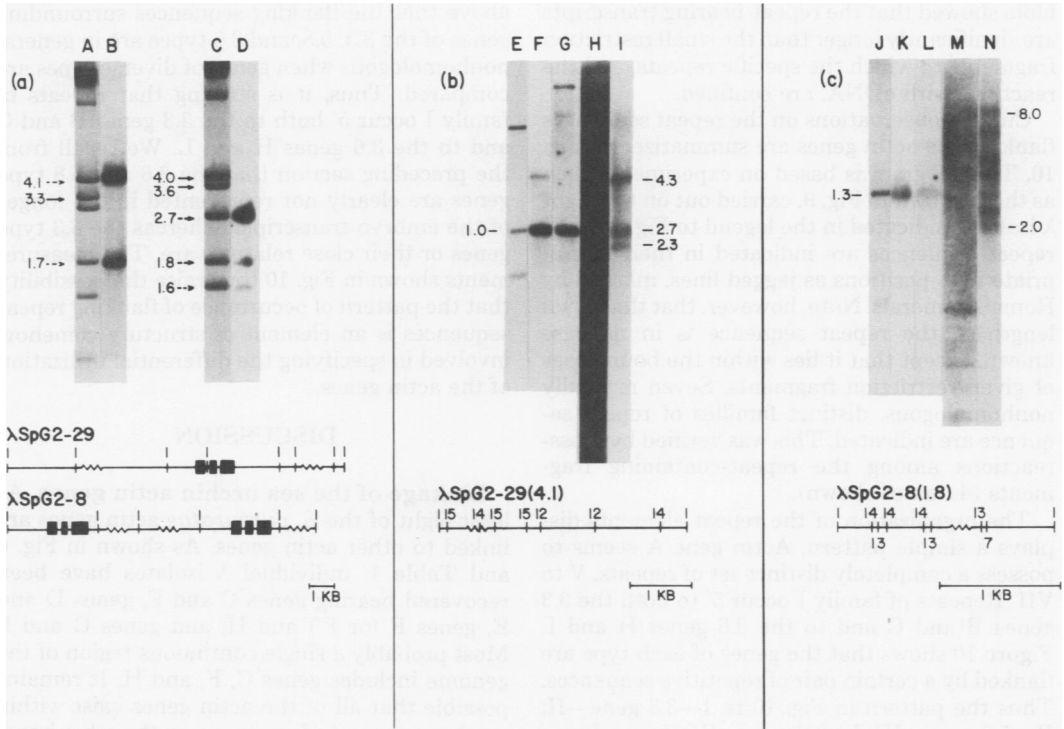


FIG. 9. Prominent repetitive sequences flanking actin genes in λ SpG2-29 and λ SpG2-8. (a) DNA of λ SpG2-29 (lane A) and of λ SpG2-8 (lane C) were digested with *EcoRI* and displayed on a gel as shown by the ethidium bromide staining patterns. These gels were blotted onto nitrocellulose filters and then hybridized with total sea urchin DNA tracer which had been labeled by nick translation to a specific activity of 5×10^6 cpm/ μ g; 4×10^6 cpm was used in each reaction. Hybridization was at 68°C in $5\times$ SET, and the filters were washed at 68°C in $1\times$ SET. The 4.1- and 1.7-kb fragments of λ SpG2-29 reacted (lane B, dashed arrows). The actin gene carried on this isolate is present on the 3.3-kb fragment (solid arrow). Similarly, the 2.7-, 1.8-, and 1.6-kb fragments of λ SpG2-8 reacted (lane D, dashed arrows), whereas the actin genes are located on the 4.0- and 3.6-kb fragments (solid arrows). The *EcoRI* maps (sites labeled 1 as previously) at the bottom of this panel are reproduced from Fig. 6 for reference. Repeats are indicated on the map by jagged lines, and actin genes are indicated by solid blocks separated by intervening sequences. (b) Further analysis of the repeat sequence in the 4.1-kb *EcoRI* fragment of λ SpG2-29 [see (a), lane B]. A map of this fragment was obtained using the enzymes 12 (*Hpa* II), 15 (*Hha* I), and 14 (*Hae* II), as shown at the bottom of the panel. The fragment was digested with *Hha*I, and the resulting subfragments are displayed in lane E. The gel was blotted and hybridized with labeled genomic DNA as in (a) (lane F); reaction occurs only with the 1-kb fragment where the repeat is located (see map). The light additional band seen in this lane is a partial digestion product since no band of this size is seen in the ethidium bromide-stained digest of lane E. In lane G the same gel blot was reacted with 10^7 cpm of [^{32}P]cDNA (10^8 cpm/ μ g) transcribed from *gastrula* poly(A) RNA. Again only the 1-kb fragment reacts. Lane H displays an *EcoRI* genome blot hybridized with the DNA of the 4.1-kb fragment labeled by nick translation, and in lane I the same tracer was reacted with a blot of a gel containing $3 \mu\text{g}$ of *gastrula* poly(A) RNA. Details of procedures for genome and RNA blots are given in the text. The transcripts observed are about 2.3, 2.7, and 4.3 kb. (c) Further analysis of the repeat sequence in the 1.8-kb *EcoRI* fragment of λ SpG2-8 [see (a), lane D]. This fragment was mapped with enzymes 7 (*Hinc*II), 13 (*Hinf*I), and 14 (*Hae*III), as shown at the bottom of the panel. Lane J shows the ethidium bromide-stained gel of a *Hinc*II digest of the fragment; lanes K and L are blots of this digest with labeled genomic DNA and cDNA, as above; lane M is an *EcoRI* genome blot in which the 1.8-kb fragment was labeled and used as probe. Comparing lane M with lane H, it can be observed that this repeat is less highly prevalent than that carried in the λ SpG2-29 fragment analyzed in (b). Lane N is a *gastrula* poly(A) RNA gel using the same probe. At least eight transcript sizes react with this repeat, ranging in size from 2 to 8 kb. Similar experiments were carried out with the 2.7-kb band of λ SpG2-8, which also contains a repeat [see (a), lane D]. However, this repeat belongs to the same family as that carried in the 4.1-kb *EcoRI* fragment of λ SpG2-29 and cross-reacts with the latter repeat. Thus the results obtained in the genome and RNA blotting experiments with the 2.7-kb band of λ SpG2-8 were identical to those shown in (b) of this figure.

blots showed that the repeat-bearing transcripts are significantly longer than the small restriction fragments to which the specific repeats, and the reactions with cDNA, are confined.

Current observations on the repeat sequences flanking the actin genes are summarized in Fig. 10. This diagram is based on experiments such as those shown in Fig. 9, carried out on the eight λ isolates indicated in the legend to Fig. 10. The repeat sequences are indicated in their appropriate map positions as jagged lines, marked by Roman numerals. Note, however, that the actual length of the repeat sequence is in no case known, except that it lies within the boundaries of given restriction fragments. Seven mutually nonhomologous, distinct families of repeat sequence are indicated. This was verified by cross-reactions among the repeat-containing fragments (data not shown).

The organization of the repeat elements displays a simple pattern. Actin gene A seems to possess a completely distinct set of repeats, V to VII. Repeats of family I occur 5' to both the 3.3 genes B and C and to the 3.6 genes H and I. Figure 10 shows that the genes of each type are flanked by a certain pair of repetitive sequences. Thus the pattern in Fig. 10 is: I—3.3 gene—II; II—5.8 gene—III; I—3.6 gene—IV. It was shown

above that the flanking sequences surrounding genes of the 3.3, 5.8, and 3.6 types are in general nonhomologous when genes of diverse types are compared. Thus, it is striking that repeats of family I occur 5' both to the 3.3 genes B and C and to the 3.6 genes H and I. We recall from the preceding section that the 3.6 and 5.8 type genes are clearly not represented in the longer of the embryo transcripts, whereas the 3.3 type genes or their close relatives are. The measurements shown in Fig. 10 thus raise the possibility that the pattern of occurrence of flanking repeat sequences is an element of structure somehow involved in specifying the differential utilization of the actin genes.

DISCUSSION

Linkage of the sea urchin actin genes. At least eight of the *S. purpuratus* actin genes are linked to other actin genes. As shown in Fig. 6 and Table 1, individual λ isolates have been recovered bearing genes C and F, genes D and E, genes F (or F') and H, and genes G and I. Most probably a single continuous region of the genome includes genes C, F, and H. It remains possible that all of the actin genes exist within one huge genomic domain, or, on the other hand,

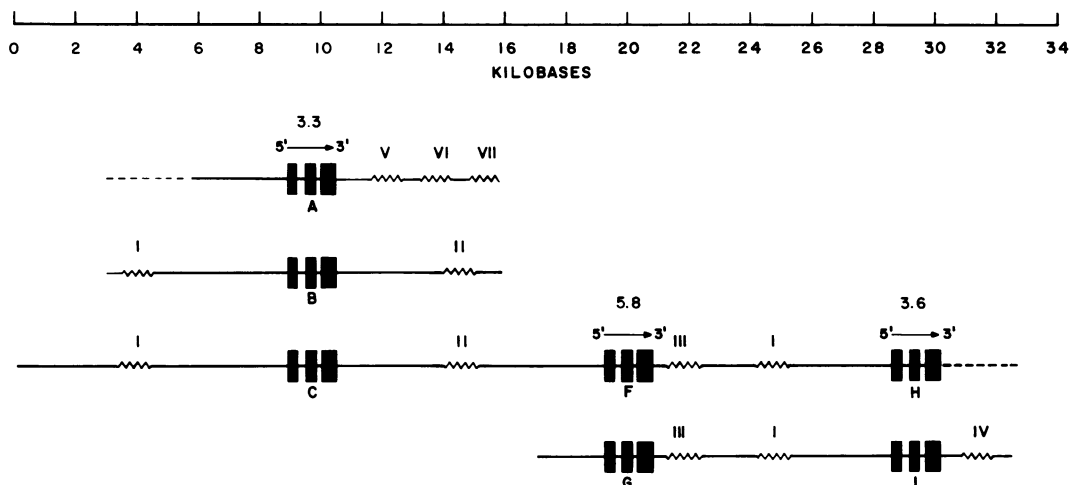


FIG. 10. Repetitive sequences flanking actin genes A to F. Repeat sequence elements of families I to VII are represented by jagged lines, and the actin gene protein coding region is indicated by solid blocks, as in earlier figures. The actin genes are designated as in Fig. 6. It is assumed that the 32-kb region containing genes C, F, and H is actually a contiguous occurrence in the genome, which is not unequivocally shown, but is probable (see text). However, no logical consequence of significance for this analysis of the pattern of repeat sequence occurrence would follow even if genes C and F were not linked directly to H, i.e., if H were linked instead to another copy (F') of the 5.8 type gene F. As indicated, genes A, B, and C are of the 3.3 type, genes F and G are of the 5.8 type, and genes H and I are of the 3.6 type. The data supporting the repeat sequence assignments shown were obtained from experiments such as those in Fig. 9, carried out on λ SpG2-29, -30, and -41 for genes C, F, and H; on λ SpG2-28 for gene A; on λ SpG2-13, -14, and -16 for gene B; and on λ SpG2-8 for genes G and I. Since the region occupied by repeat I is not included on λ SpG2-28, a dashed line is shown in place of repeat I to the left of gene A.

TABLE 1. Summary of relationships among 11 cloned actin genes

Actin gene	Type designation	Share 3' SpG2 sequence	Share 3' gene F sequence	Intervening sequence homology with gene C	Intervening sequence homology with gene H	Linkage
A	3.3	+	-	+	-	
B	3.3	+	-	+	-	
C	3.3	+	-	+	-	F
D	Related to 3.3	+	-			E
E	Related to 3.3	+	-			D
J	SpG28	+	-			
F	5.8	-	+	+	-	C, H
G	5.8	-	+	+/-	-	I
H	3.6	-	+	-	+	F
I	3.6	-	+	-	+	G
K	11.0	-				

that they occur in a number of distant linkage groups. Since the distance separating actin genes which are known to be linked is more than 50% of the average length of the inserts in the λ recombinants, there is a good probability of not recovering two genes on a single isolate even if they are linked at the usual distances in the genome. The arrangement observed for the actin genes of the sea urchin is clearly different from that seen in *Drosophila*, where all six actin genes occur singly in different genomic locations (13a, 14, 34).

All actin gene pairs so far known are heterotypic, that is, the linked genes are not of the same type even though multiple copies of these types exist in the genome (D and E could represent a possible exception). Thus the linkages observed associate a 3.3 type gene, C, with a 5.8 type gene, F, whereas the 3.6 type genes H and I are linked, respectively, to the 5.8 type genes F and G. Heterotypic linkage of genes coding for functionally related proteins is known in other systems as well, e.g., the histone genes of the sea urchin (4), the chorion protein genes of wild silk moths (17) and *Drosophila* (33), and the globin genes of mammals (13, 20, 21). In the globin gene system, heterotypic linked genes are expressed at different stages and in different cells, whereas in all the other examples cited the linked genes appear to function at more or less the same time in the same differentiated cells. This has given rise to the reasonable speculation that such coordinately active linked genes belong to defined regions of the genome which act as local regulatory domains. However, it is likely that the organization of the sea urchin actin genes is of the globin type. Thus gene F, which is known not to be represented by a predominant form of maternal message, is linked to gene C, which is of a type that is represented by a 2.2-kb egg transcript. It will be interesting (though difficult) to determine whether gene C itself is transcribed during oogenesis, or only other 3.3 type genes

which lie in separate locations.

The sea urchin actin gene types and their evolution. Until actin genes of the several types defined here are sequenced, we will not know whether they actually code for significantly different proteins. We are unaware of any sequence data demonstrating heterogeneity of sea urchin actin proteins, though electrophoretic variants have been noticed (11, 16). However, unpublished partial sequence data of Crain and his associates show that both 3' and 5' coding regions of the gene carried in SpG28 (gene J) differ by 12 to 15% in sequence from the equivalent regions of the 3.3 type gene carried in SpG17 (i.e., gene C). Most of the coding regions of the various actin genes shown in Fig. 6 differ, if at all, to a much lesser extent. Thus, in experiments not shown here, we measured the thermal stability of heteroduplexes between the coding regions of actin genes representative of the 3.3, 3.6, and 5.8 types and found them all to be identical within the limits of resolution of the experiment, which was about 2%. This amount of variation could of course be highly significant for function since vertebrate actins found in different tissues often vary by only a small number of amino acids (36).

The elements of the actin genes distinguishing the diverse types are the intervening sequences (Fig. 3 and 7), the portions giving rise to 3' untranslated message sequences (Fig. 4 and 5), and the flanking sequences on both 5' and 3' ends which are not included in the mature messages (Fig. 3, 6, and 7). These distinctions are summarized in Table 1. The actin genes seem to fall into two general classes, distinguished by the 3' sequences abutting their coding regions. Differences among the 3'-terminal regions of multiple actin genes are apparently common. Such differences have been reported in chickens (29), in *Drosophila* (13a), and in *Dictyostelium* (28). An interesting possibility relevant to the two sea urchin embryo transcripts is that the 3' untrans-

lated tail of the 2.2-kb transcript confers on it an increased metabolic stability. This proposition is suggested by the observation that the 2.2-kb transcript is the main form observed in the egg (Fig. 8). A general characteristic of maternal mRNA's is their great metabolic stability during oogenesis (7), and some egg messages are now known to persist well into embryogenesis. A direct measurement, although preliminary, indicates that the 2.2-kb transcript is very stable in gastrula stage embryos, whereas the 1.8-kb transcript turns over with about a 3-h half-life (C. V. Cabrera, J. J. Lee, and E. H. Davidson, unpublished data).

Intervening sequences that occur in the same position in the coding sequences of evolutionarily related genes, but are at least partially non-homologous, have been observed in a number of gene systems. Among the most striking examples are the vitellogenin genes (37), the globin genes (23), and the ovalbumin-X-Y gene set (30). Obviously the presence of intervening sequences and their locations are relatively conservative evolutionary features, whereas most of their internal sequence is not. In all of the sea urchin actin genes where the positions of the intervening sequences are known (A to C, F to I), the coding region is apparently interrupted at the same places in the amino acid sequence. These interruptions are known to occur at positions 121/122 and 203/204 in gene C (11). One possibility is that these two intervening sequences were present in the common ancestor of all of these actin genes, whereas the intrusion of the additional two intervening sequences at positions 41/42 and 267/268 in gene J (Crain et al., unpublished data) was a later event. The location of none of the known intervening sequences in the sea urchin actin genes conforms to their position in actin genes of the other organisms for which data are available. In *Dictyostelium* no actin intervening sequences are yet known (12); in the yeast actin gene there is an intervening sequence at position 3 in the amino acid coding sequence; in *Drosophila* there are intervening sequences at amino acid positions 14 and 307 (13a). This suggests that though the actin gene itself is very old, the evolutionary gene duplication process which gave rise to the *S. purpuratus* actin gene set occurred since the divergence of the deuterostomes, and perhaps even within the radiation of the echinoderms.

Many systems of evolutionarily related multiple genes are now known. Although it is not yet proven that the following is true for the sea urchin actin genes, it is often the case that particular types or copies of related multiple genes are utilized in specific states of differentiation. Such multiple genes have apparently arisen

by ancient evolutionary duplications and since their separation have diverged to some extent in sequence and structure. The most important distinction between these gene cousins, however, is that they belong to different regulatory pathways. This argument may of course be inverted. Thus we propose that in evolution a new state of differentiation (e.g., a new cell type, a new developmental pathway, or a new tissue specialization) could be caused by the duplication of one or several genes and their insertion into a genomic domain already involved in specification of an extant state of differentiation or developmental structure. In other words, the key event envisioned is the dispersal of new copies of previously existent genes into different developmental regulatory units. Thereafter the duplicated genes will diverge along separate pathways and will come to differ according to chance events and to the different selective processes resulting from their function in entirely new biological contexts. This argument suggests that evolutionary transformation of preexisting structures in new and diverse directions will often be associated with duplication of genes or of whole multigenic domains; that many structural genes will be found which are evolutionarily related but function in different cell types; and also that the particular duplication events leading to the sets of genes present will often be entirely specific to various branches of evolution. That is, we imply that the formation of small families of related genes functioning in different cell types is the event which results in the changes in organism structure which we recognize as defining phylogenetic groups.

Repetitive sequences around the sea urchin actin genes. The distribution of repeat sequences around the actin genes is correlated to a large extent with the gene type, as shown in Fig. 10. However, among different copies of given actin gene types the overall flanking sequence regions tend to be similar, so it is not surprising that the same is true of the repetitive sequences embedded in these regions. Thus the patterns of repeat sequence distribution we observe could simply reflect the series of evolutionary DNA duplications by which the multiple actin genes were created. Some difficulty for this explanation may be produced by the observation of elements of repeat family I to the 5' side of both the 3.3 type genes B and C and the 3.6 type genes H and I in an otherwise nonhomologous environment. Irrespective of the origin of the repeat sequence patterns, however, structural studies such as these can be at best merely suggestive. The essential issue they raise is that of the flanking repetitive sequence functions, if any.

The role of repetitive sequences in genomic expression has so far resisted elucidation. One important observation, which in our opinion suggests there is such a role, is the widespread but tissue- and stage-specific transcription of many species of repeat sequence into heterogeneous nuclear RNAs and also into egg poly(A) RNA in sea urchins (5, 6, 31). The repeats surrounding the actin genes are typical in that they belong to families represented by relatively long (i.e., several kb) embryo poly(A) RNA transcripts. However, these particular repeat sequences are not detected in stable transcripts in embryos, whereas at least several other members of the same repeat families are (Fig. 9). It is of course possible that some or all of the repeats flanking the actin genes are indeed transcribed into unstable nuclear RNAs, e.g., actin mRNA precursors. Such transcription could be related (positively or negatively) to actin gene expression. Each of the flanking repeat sequence families represents a network of homology which associates the actin genes with other regions of the genome. These networks are different for the various actin gene types and provide the possibility for specific interactions with distant homologs at either RNA or genomic levels.

ACKNOWLEDGMENTS

We are grateful to many of our associates and colleagues for assistance, criticism, and insight at many stages in this work. Laurence Lasky carried out some of the initial screening and blot hybridization experiments which drew attention to SpG2 as a clone of interest. David Anderson isolated and mapped several of the actin λ isolates and produced preliminary evidence that mRNA's coding for actin protein are selected from embryo poly(A) RNA by SpG2. M. Chamberlin assisted in the completion of some key experiments, and Barbara Hough-Evans provided a most useful critical review of a draft of the manuscript. Special thanks are also due to Norman Davidson and Richard A. Firtel for interesting and perspicacious reviews as well.

This research was supported by Public Health Service grants GM-20927 (to E.H.D. and R.J.B.) and GM-24620, GM-25492, CA-12708, and RR-05528 (to W.R.C.) from the National Institutes of Health. R.H.S. was supported by Public Health Service postdoctoral training grant GM-07401, and J.W.P. was supported by Public Health Service predoctoral training grant GM-07616, both from the National Institutes of Health. L.B.M. received support from a California Institute of Technology Summer Undergraduate Research Fellowship. R.J.B. is also a staff member, Carnegie Institution of Washington, Washington, D.C. D.S.D. was supported by a fellowship from the Muscular Dystrophy Foundation.

LITERATURE CITED

- Anderson, D. M., R. H. Scheller, J. W. Posakony, L. B. McAllister, S. G. Trabert, C. Beall, R. J. Britten, and E. H. Davidson. 1981. Repetitive sequences in the sea urchin genome. Distribution of members of specific repetitive families. *J. Mol. Biol.* 145:5-28.
- Brosius, J., T. J. Dull, and H. F. Noller. 1980. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 77:201-204.
- Brosius, J., M. L. Palmer, P. J. Kennedy, and H. F. Noller. 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 75:4801-4805.
- Cohn, R. H., J. C. Lowry, and L. H. Kedes. 1976. Histone genes of the sea urchin (*S. purpuratus*) cloned in *E. coli*: order, polarity, and strandedness of the five histone-coding and spacer regions. *Cell* 9:147-161.
- Costantini, F. D., R. J. Britten, and E. H. Davidson. 1980. Message sequences and short repetitive sequences are interspersed in sea urchin egg poly(A)⁺ RNA. *Nature (London)* 287:111-117.
- Costantini, F. D., R. H. Scheller, R. J. Britten, and E. H. Davidson. 1978. Repetitive sequence transcripts in the mature sea urchin oocyte. *Cell* 15:173-187.
- Davidson, E. H. 1976. Gene activity in early development. Academic Press, New York.
- Davidson, E. H., B. R. Hough, W. H. Klein, and R. J. Britten. 1975. Structural genes adjacent to interspersed repetitive DNA sequences. *Cell* 4:217-238.
- Davis, R. W., M. Simon, and N. Davidson. 1971. Electron microscope heteroduplex methods for mapping regions of base sequence homology in nucleic acids. *Methods Enzymol.* 21D:413-428.
- Denhardt, D. T. 1966. A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* 23:641-646.
- Durica, D. S., J. A. Schloss, and W. R. Crain, Jr. 1980. Organization of actin gene sequences in the sea urchin: molecular cloning of an intron-containing DNA sequence coding for a cytoplasmic actin. *Proc. Natl. Acad. Sci. U.S.A.* 77:5683-5687.
- Firtel, R. A., R. Timm, A. R. Kimmel, and M. McKeown. 1979. Unusual nucleotide sequences at the 5' end of actin genes in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. U.S.A.* 76:6206-6210.
- Fritsch, E. R., R. M. Lawn, and T. Maniatis. 1980. Molecular cloning and characterization of the human β -like globin gene cluster. *Cell* 19:959-972.
- Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixer, and N. Davidson. 1981. The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. *Cell* 24:107-116.
- Fyrberg, E. A., K. L. Kindle, and N. Davidson. 1980. The actin genes of *Drosophila*: a dispersed multigene family. *Cell* 19:365-378.
- Graham, D. E., B. R. Neufeld, E. H. Davidson, and R. J. Britten. 1974. Interspersion of repetitive and non-repetitive DNA sequences in the sea urchin genome. *Cell* 1:127-137.
- Infante, A. A., and L. J. Heilmann. 1981. Distribution of messenger ribonucleic acid in polysomes and nonpolysomal particles of sea urchin embryos: translational control of actin synthesis. *Biochemistry* 20:1-8.
- Jones, C. W., and F. C. Kafatos. 1980. Coordinately expressed members of two chorion multigene families are clustered, alternating and divergently orientated. *Nature (London)* 284:635-638.
- Kaback, D. B., L. M. Angerer, and N. Davidson. 1979. Improved methods for the formation and stabilization of R-loops. *Nucleic Acids Res.* 6:2499-2517.
- Kindle, K. L., and R. A. Firtel. 1978. Identification and analysis of *Dictyostelium* actin genes, a family of moderately repeated genes. *Cell* 15:763-778.
- Konkel, D. A., J. V. Maizel, Jr., and P. Leder. 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β -globin genes. *Cell* 18:865-873.
- Lacy, E., R. C. Hardison, D. Quon, and T. Maniatis. 1979. The linkage arrangement of four rabbit β -like globin genes. *Cell* 18:1273-1283.

22. Lasky, L. A., Z. Lev, J.-H. Xin, R. J. Britten, and E. H. Davidson. 1980. Messenger RNA prevalence in sea urchin embryos measured with cloned cDNAs. *Proc. Natl. Acad. Sci. U.S.A.* 77:5317-5321.
23. Leder, P., J. N. Hansen, D. Konkel, A. Leder, Y. Nishioka, and C. Talkington. 1980. Mouse globin system: a functional and evolutionary analysis. *Science* 209:1336-1342.
24. Lee, A. S., T. L. Thomas, Z. Lev, R. J. Britten, and E. H. Davidson. 1980. Four sizes of transcript produced by a single sea urchin gene expressed in early embryos. *Proc. Natl. Acad. Sci. U.S.A.* 77:3259-3263.
25. Lev, Z., T. L. Thomas, A. S. Lee, R. C. Angerer, R. J. Britten, and E. H. Davidson. 1980. Developmental expression of two cloned sequences coding for rare sea urchin embryo messages. *Dev. Biol.* 76:322-340.
26. Maniatis, T., R. C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Kim, and A. Efstratiadis. 1978. The isolation of structural genes from libraries of eucaryote DNA. *Cell* 15:687-701.
27. Maxam, A. M., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. U.S.A.* 74:560-564.
28. McKeown, M., W. C. Taylor, K. L. Kindle, R. A. Firtel, W. Bender, and N. Davidson. 1978. Multiple, heterogeneous actin genes in *Dictyostelium*. *Cell* 15:789-800.
29. Ordahl, C. P., S. M. Tilghman, C. Ovitt, J. Fornwald, and M. T. Largen. 1980. Structure and developmental expression of the chick α -actin gene. *Nucleic Acids Res.* 8:4989-5005.
30. Royal, A., A. Garapin, B. Cami, F. Perin, J. L. Mandel, M. LeMeur, F. Brégégère, F. Gannon, J. P. LePenec, P. Chambon, and P. Kourilsky. 1979. The ovalbumin gene region: common features in the organization of three genes expressed in chicken oviduct under hormonal control. *Nature (London)* 279:125-132.
31. Scheller, R. H., F. D. Costantini, M. R. Kozlowski, R. J. Britten, and E. H. Davidson. 1978. Representation of cloned interspersed repetitive sequences in sea urchin RNAs. *Cell* 15:189-203.
32. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
33. Spradling, A. C., M. E. Digan, A. P. Mahowald, M. Scott, and E. A. Craig. 1980. Two clusters of genes for major chorion proteins of *Drosophila melanogaster*. *Cell* 19:905-914.
34. Tobin, S. L., E. Zulauf, F. Sanchez, E. A. Craig, and B. J. McCarthy. 1980. Multiple actin-related sequences in the *Drosophila melanogaster* genome. *Cell* 19:121-131.
35. Vandekerckhove, J., and K. Weber. 1978. Actin amino-acid sequences. Comparison of actins from calf thymus, bovine brain, and SV40-transformed mouse 3T3 cells with rabbit skeletal muscle actin. *Eur. J. Biochem.* 90:451-462.
36. Vandekerckhove, J., and K. Weber. 1978. At least six different actins are expressed in a higher mammal: an analysis based on the amino acid sequence of the amino-terminal tryptic peptide. *J. Mol. Biol.* 126:783-802.
37. Wahli, W., I. B. Dawid, T. Wyler, R. Weber, and G. U. Ryffel. 1980. Comparative analysis of the structural organization of two closely related vitellogenin genes in *X. laevis*. *Cell* 20:107-117.
38. Zain, S., J. Sambrook, R. J. Roberts, W. Keller, M. Fried, and A. R. Dunn. 1979. Nucleotide sequence analysis of the leader segments in a cloned copy of adenovirus 2 fiber mRNA. *Cell* 16:851-861.